MACHINE LEARNING:

The key to understanding the decision making of the US Supreme Court?

Author

Alice Rázková

Supervisor

Andrej Savin

Co-Supervisor

Dolores Romero Morales

A thesis submitted for the partial fulfilment
of the requirements for
the degree of

MASTER OF SCIENCE

in

Business Administration and Information Systems (e-Business)

Submitted:	1st December 2016
Number of pages:	56
Character Count:	131.258

Abstract

The US Supreme Court enjoys power and authority unseen in any other jurisdiction in the world, which causes speculation about the basis for its decision-making and the factors influencing the decisions arrived at by the Court.

Its structure, while ostensibly demonstrating clear independence, can be subject to misuse. Its decision making process is so complex that a human being, even an expert, can have difficulties understanding it.

The Supreme Court provides thousands of documents containing information about each of the cases it decides upon and other relevant information, which can be analysed with machine learning methods to provide insight into the patterns, outliers and the most significant factors in the decision making of the justices, which helps to more fully understand their decision making.

This thesis seeks to remove the ‘blackbox’ of decision making of the US Supreme Court, and to present Machine Learning not only as a tool to understand the Court, but also as a solution to monitor potential bias in its decisions. This is a great opportunity for businesses, where Supreme Court rulings can have a major impact on their bottom line, to better help them assess every potential threat and opportunity.

Table of contents

1) Introduction

a) Context

The US Supreme Court by design keeps its operations largely secret from the outside world. Not surprisingly, its decisions attract the attention of businesses, the press and the people. The decision making of the Court is one of the biggest question marks for those it affects and for anyone who seeks to understand, analyse and predict its outcomes, which by its very definition have far reaching effects.

The character of the US Supreme Court is perfectly reflected by the structure of its building, designed by Cass Gilbert with a massive and imposing stairway. According to Jeffrey Toobin, the walk up the stairs would be the central symbolic experience of the Supreme Court, the physical manifestation of the American march to Justice. This stairway would separate the Court from the rest of the world, announcing that the Justices operate on a higher plane, which seems like the perfect physical representation of the Court's status. It is the highest instance in the judicial review, which gives the Court the power to rule on constitutionality of laws, defining what law is and what it is not.

The secrecy around the Supreme Court is not very comforting for those whose lives depend on its decisions and who have to live with its decisions, often without the chance for another appeal (Vernon Hugh Bowman v. Monsanto Company, et al., docket no.: 11-796). This lack of transparency has resulted in a lack of trust in the Supreme Court causing doubts about its legitimacy. (The New York Times, 2012), (Posner E. , 2015), (McCarthy, 2015). Currently it is seen more as a political institution than as an impartial Court. (Greenhouse, 2016)

It is therefore important to find a way to regain confidence in the Supreme Court by improving our understanding of the way the Court makes its decisions and to clarify what the most important factors are in such decision making.

Empirical evidence, showing that the Supreme Court Justices are not acting solely on their political preferences but rather only base their judgements on their own understanding of the constitution would not only improve the lost confidence in the Supreme Court, but also enhance understanding of its decision making process.

A lot of developments have taken place in Machine Learning which continue to further our understanding of nearly every aspect of our lives. With Moore's law, we get access to an increasing amount of data for a constantly decreasing price, which opens up endless possibilities to extract actionable insights from collected data points, transforming the data into information, giving us the necessary information for informed decision making.

In the late 1980s, Harold J. Spaeth provided significant insights into the cases and political factors that helped drive the Court's decision making by publishing a database of classified cases with every single vote by a Supreme Court Justice in all argued cases over a five-decade period (Law, The Genesis of the Database). Despite the publication of the database, little further work was carried out in this area in the following 20 years. Finally, in 2002, a forecasting project by Andrew Martin, Kevin Quinn, Theodore Ruger and Pauline Kim, where the machines and legal experts were predicting the outcomes of the US Supreme Court, showed that machines outperformed human beings. (Martin, Quinn, Ruger, & Kim, 2004)

Since then, Machine Learning has found many more applications in the legal sector, to name but a few, in discovery, compliance, and in contracts. Advanced analytics have been changing how lawyers find their clients, conduct legal research, how discovery is conducted and how contracts are drafted and verified. It is changing not only billing but also performance management, selection of juries and prediction of case outcomes. According to a survey conducted in March 2015, 90% of corporate legal departments, law firms and government lawyers are using data analytics, naming cost management and savings as one of the most important benefits, and almost a third of respondents cited improved decision making. (Gabe, Survey: Some Lawyers View Data As Threat, Investment Growing, 2015)

The market for legal analytics is growing slowly due to the legal profession being traditionally one of the most conservative academic fields, however according to the CEO of Ravel Law (providing judge analytics), *"(the legal sector) has become a lot more receptive and interested in technology over the past several years".* (Marr, 2016) The above mentioned 2015 survey also shows the willingness of the legal sector to extend the use of advanced analytics as 68% of the respondents expected legal spending to increase in the following 2 years. (Gabe, Survey: Some Lawyers View Data As Threat, Investment Growing, 2015)

The reason, why this change is moving at a faster pace than usual are the advantages of advanced analytics. These include more space, as there is no longer a need for storage rooms full of case related information in paper files, which is now available online and already filtered to ease the administrative burden.

Not only does it save space and time one would spend looking for particular information due to the amount of files an algorithm can process, but it can also find information one would miss or would not be able to find in a short period of time. This includes for example unexpected judgements or decisions, helping a relevant case.

Further, it improves transparency and efficiency, through better organisation, taking a data driven approach, and by removing human related bias.

Last but not least, it leaves more time for work which cannot be automated (yet), such as litigation strategy or consultation with clients, reducing billable hours, promoting access to justice and at the same time, reducing the amount of errors caused by human bias, fatigue or simply lack of information, which may result in cases ending up in the court of appeals or retrials.

This particular research is based on a prediction model created by Daniel Martin Katz, Michael J Bommarito II, and Josh Blackman, which makes predictions about the outcomes of the US Supreme Court based on publicly available data, and brings more clarity into the aspects influencing the decisions of the US Supreme Court. Moreover, it gives an additional piece of information to astute managers who may seek to use the law to their advantage. The model correctly identified 69.7% of the Court's overall affirm/reverse decisions and correctly forecasted about 70.9% of the votes of individual justices over a time span of 60 years. Due to its robust as well as generalized and fully predictive character, it can predict outcomes even when the composition of the Court changes. It also identified the most important factors in the decision making, demonstrating that personal beliefs do not play the most important role.

This model has been chosen as it is so far the best in existence at predicting the outcomes and in bringing transparency into the decisions of the US Supreme Court and that of any case decided by the Court since 1953.

Although this prediction model is predicting the outcomes of the US Supreme Court, it does not exclude the possibility of creating a model predicting court decisions outside the US jurisdiction. In fact, on October 24th 2016, researchers at University College London, University of Sheffield and the University of Pennsylvania, developed a similar prediction model, which with the help of Machine Learning, has been shown to predict the judicial decisions of the European Court of Human Rights to 79% accuracy. It is the first to predict the outcomes of a major international court by analysing case text. (Aletras, Tsarapatsanic, Preoţiuc-Pietro, & Lampo, 2016)

b) Motivation

The understanding of the decision making of the US Supreme Court is based on a series of assumptions and assertions.

> *"It is commonly assumed that Supreme Court justices' votes largely reflect their attitudes, values, or personal policy preference. Nevertheless, this assumption has never been adequately tested with independent measures of the ideological values of justices, that is, measures not taken from their votes on the Court"* (Segal & Cover, 1989)

The technology of today provides tools to tackle the information problem and replace the assumptions with facts deduced from collected data. Machine Learning allows for data-driven decisions, which are increasingly favoured in many sectors, including medicine, where decision making is facilitated by Machine Learning algorithms which screen all data relevant to each patient (Chi, 2009), (Judex.com, 2016). The doctors then only check the choice for irregularities. Even the business sector has started taking data-driven decision making seriously, as it allows for better interpretation of each decision and prediction of future events. Moreover it proved to have a positive effect on performance and efficiency (Brynjolfsson, Hitt, & Kim, 2016), (Reeves & Ueda, 2016).

This thesis hopes to endorse the same technique in the legal sector and its business applications, more specifically in the context of Supreme Court decision making, bring clarity, better performance of the Court, the missing review of the Court as well as business implications. Additionally, it wants to make recommendations to the current prediction model that would make the model in the case study more relevant for businesses.

Prediction models are just the first step in understanding the decision making of the US Supreme Court, which will further improve the confidence in the judicial branch. By reducing uncertainty, it will improve the process of decision making by businesses, drive down litigation costs and bring much-needed competition into the legal sector.

c) Delimitation

This section aims to explain the choices made in connection with the scope of this thesis.

The thesis focuses on the use of Machine Learning in the legal sector in the United States, and the only applications currently in use in businesses being used on the US market. Another reason for the limitation is the legal structure of the United States. The US Supreme Court enjoys unrivalled power, which has an impact on the rest of the judicial hierarchy of the United States.

The thesis also eliminates any cases not involving business for the aim of this thesis is to help businesses to make better informed decisions. The direct factors in these decisions are cases involving contract law, licensing, e-commerce, torts and privacy litigation, product liability, intellectual property, employment agreement, employer discrimination, environmental law, international law and law of transactions.

Further, it is believed that the better the data the better the predictions. In the context of Machine Learning, the timing becomes important because the data of today is more similar to the data of tomorrow than the data of yesterday. The more accurate the data, the more accurate the results will be, which is why only the cases under the current Chief Justice, Roberts, are taken into account.

The methodological perspective reflects the author’s beliefs and worldview, but was also the best option, given the multi-disciplinary approach, covering not only law, psychology but also a scientific discipline – Machine Learning.

d) Research question

The research question of this thesis aims to study the advantages of Machine Learning in the legal sector, more specifically, how can Machine Learning help the understanding of the process of decision making by the US Supreme Court? How can the prediction model by Katz, Bommarito II, & Blackman be optimised to provide better insights for businesses and what are the advantages of such prediction models?

e) Readers' guide

Introduction	Introduces the topic of the research with questions the research seeks to answer, gives clear objectives, author's motivations and delimitations of the scope of the research.
Methodology	Presents the theoretical and philosophical assumptions upon which this research was undertaken and its influence on the research, particularly the reason for using three disciplines, and the effects of critical realism on the deductive and inductive research approach chosen in this research.
Decision making: Connection Between Law and Machine Learning	This chapter comprises three different topics, identified as essential for understanding of the conducted research. They clarify what law and Machine Learning have in common and why adopting Machine Learning in the legal sector is not an option but a must.
What can a Machine Learning Algorithm tell us about the decision making of the US Supreme Court?	This chapter discusses the qualities of a Machine Learning model in connection with the US Supreme Court.
Why do we need Machine Learning on the Court?	In this chapter, the author takes into account all problems that the decision making of the US Supreme Court is facing and presents a solution for the increasing amount of information as well as an increasing uncertainty.
How can the understanding of the US Supreme Court be further improved?	This chapter focuses on possible improvements of the existing prediction model. It includes hypothesis about possible improvements in accuracy and its validation.
What Does Machine Learning Mean for the Court and for Businesses?	In this chapter the author seeks to summarize the needs and opportunities for the Court and for businesses, which rely on the outcomes.
Concluding Remarks	The last part of this thesis summarizes the findings and makes recommendations for future research.

2) Methodology

"Content without methods leads to fantasy; method without content to empty sophistry" (Johann Wolfgang Goethe, 1892)

To avoid the former scenario, a methodological framework was designed to support the qualities of this thesis, mainly, its multidisciplinary character, which was chosen to get a better perspective on the issue.

This thesis relies on a number of disciplines, which are crucial to the understanding of the research question and herewith for the answer of the research question.

The first discipline is the legal discipline, which provides the context and the rules governing the US Supreme Court. It explains the challenges of the political structure of the Government, which is causing people to doubt the Supreme Court's trust and authority.

The second discipline is psychology, more specifically in decision-making, which by itself already relies on a number of disciplines - it is therefore not surprising to see it in a multidisciplinary context.

The third discipline is Machine Learning, which provides a solution for the legal challenges, combined with psychological challenges.

All three disciplines are necessary for this research as they all contribute equally to the understanding of the topic. Without the legal discipline, the understanding of legal challenges would not be clear. Without the understanding of the psychology of human decision making, the cognitive limitations would not be clear and only half of the problem would have been addressed. Without the Machine Learning discipline, the solutions would have remained hypothetical as no context would have been provided.

Although this thesis recognizes that all three approaches cannot be sufficient to answer the question from all angles, the multidisciplinary approach gets closer to the big picture. A clear weakness of this approach is the necessary restriction of depth in each topic. However, by bringing in the key insights of each of the disciplines involved, it provides a broader picture and a better understanding of the topic and its context.

By using them all together, the most important limitation – the lack of perspective - is reduced, allowing for a more rounded understanding of the problem area. Relying solely on one discipline would not cover the research question, which already mentions law, Machine Learning as well as decision making. In order to answer the research question, all of the elements in the question have to be understood, in order to get an idea about the far-reaching advantages of a machine learning approach.

This thesis has legal context as its backbone, but it relies equally on decision making and machine learning methods, which are in the immediate context. Not understanding the legal context would result in lack of understanding of why the algorithm is needed and what it is supposed to achieve. The psychological approach then clarifies human limitations, which are to be addressed in decision making, and the Machine learning approach then takes all the information into consideration, creating an algorithm, that increases understanding, which serves as further review of the Supreme Court, while using the information available to legal experts.

This is represented in the structure of the thesis, which starts with legal context, continues with decision making and ends with machine learning techniques. All three disciplines are then combined to provide an idea about business application.

It is important to note that while the main research question is *how Machine Learning can improve the understanding of the decision making of the US Supreme Court*, other stakeholders need to be taken into account too. Various stakeholders influence the decision making of the US Supreme Court. At the same time, the decision making influences other parts of the environment, and the thesis will thus address also these stakeholders, which are influenced by the US Supreme Court.

This methodological design is carefully drafted to support this multidisciplinary approach, by applying critical realism as a philosophical approach, which stresses the importance of meaning as well as context surrounding the area being researched.

Further, by using both deductive and inductive methods it provides a strong backbone of theory coming from the reviewed literature, while at the same time providing freedom to reformulate ideas and evolve as new observations are made and the context becomes clear.

This research applied a mixed-model strategy, which gives the researcher even more freedom by removing the limitation to use only one source of data. Qualitative and quantitative data are equally important to the research, because they bring more clarity about the context and put numbers behind the theory.

The following section will make clear, why these choices are indeed meaningful.

i) Ontological and epistemological

Considering the context and topic in the drawing up of this research, the realism approach was chosen to analyse the field of research and to apply the right methodology to carry out it out.

There are two additional approaches to realism, namely direct realism and critical realism. The direct realism approach states that "What you see is what you get", i.e. that which is experienced in the senses are also portrayed accurately in the world. (Saunders, Thornhill, & Lewis, 2009). However, this research will use the critical realism approach, since the ontology of critical realism states that: "(…) there is a reality independent of our thinking about it that science can study" (Trochim, 2016)

This ontological view acknowledges that knowledge exists independently of human thoughts and beliefs of the actors (Saunders, Thornhill, & Lewis, 2009). These thoughts and beliefs are interpreted through social conditioning, which further entails that the personal views of the actors in the research can impact the results.

Critical realism further implies that the researcher is biased by worldviews and cultural experiences, which impact the research (Saunders, Thornhill, & Lewis, 2009). The author will try to understand the external reality and describe the structures and connections between the social actors. It implies that the researcher acknowledges that there exists social connections that cannot be described, but which are impacting, and have the ability to change, the social objects even though they are not directly quantifiable. (Trochim, 2016)

From an epistemological consideration, this further applies that phenomena can be misinterpreted and that it is important that the context of this study also acknowledges that the conclusions derived can vary, if carried out by other researchers with other ontological and epistemological definitions.

The author pursues critical realism, because its ontology is the attempt to understand and say something about the things themselves and not about our beliefs, experiences, of our current knowledge and understanding of those things. Critical realism argues that to understand the reality, uncovered by science and social science we need a structured and differentiated account, in which openness, difference, stratification and change is central. *As such, critical realism is epistemologically focused on credible data and explaining phenomena within a context or contexts.*

Credible data can enable understanding of specific observations, however the sensations created by these observations still risk being mis- and reinterpreted. (Saunders, Thornhill, & Lewis, 2009)

Ontologically, critical realism lends itself to an objective perspective. The world does indeed exist independently of human thought, but is interpreted through critical thinking and social conditioning (Saunders, Thornhill, & Lewis, 2009). Unlike social constructivists, critical realists accept that a reality exists beyond those social actors, and that one cannot focus purely on the observable reality or the social constructs that surround it exclusively.

Knowledge acquisition does not happen in a continuous nor in a discontinuous process through changes. Although our reality does not change, our understanding of it changes due to fluctuating perception of our social structures. The difficulty with critical realism is that we can never fully know. While changes in data can influence our understanding of reality, so can changes in interpretation. This thesis lends itself to a critical realist methodological philosophy for many different reasons.

First of all, the focus on meaning, interpretation, and context supports a multi-disciplinary approach. While the main discipline of this thesis is rooted in decision making, this thesis recognizes that various disciplines can offer different contextual insights into the research question. Critical realism endorses this identity, claiming that what we observe is only part of the bigger picture (Saunders, Thornhill, & Lewis, 2009).

Data driven decision making has an impact on different parties. Allowing for a multi-level approach would therefore support the methodology of critical realism by discovering various contextual information relevant to the research question.

From an axiological perspective, the critical realist is subject to biases that influence the research and therefore the outcome. Upbringing, cultural experience, and world views are typical sources of biases, which influence the researcher. To minimize the bias, it is important to get a sense of context through reviewing different sources of literature, which not only improves the validity of the research through increased objectivity, but also improves the results through discovery of new elements relevant to the research.

Critical realism allows for a multivariate data collection, which is very important for this research, given it uses qualitative as well as quantitative approaches to data. By using both approaches, the researcher gets a fuller picture of the issue and herewith better accuracy of the results.

Unlike constructivism which relies heavily on qualitative data, and positivism which on the other hand relies on quantitative data, creating a clear bias, the methodology used in this thesis – critical realism - gives these two sources equal weight and uses the qualities of both approaches to achieve better validity of the research.

ii) Research Strategy

According to Cohen and Kaplan, an applied theory is never considered complete, rather it is *"true until shown otherwise"* (Kaplan, 1964), (Cohen, 1991). Some researches begin with deduction, and at some point they become informed by induction. Other researches may develop the other way around, it all depends on the specific nature of the phenomenon the research focuses on, not on the researchers preferences as some believe. According to Susan A. Lynham, each specific method of applied theory building is a way of developing insight, understanding, and possible explanation of the phenomenon, issue, or problem. (Lynham, 2016)

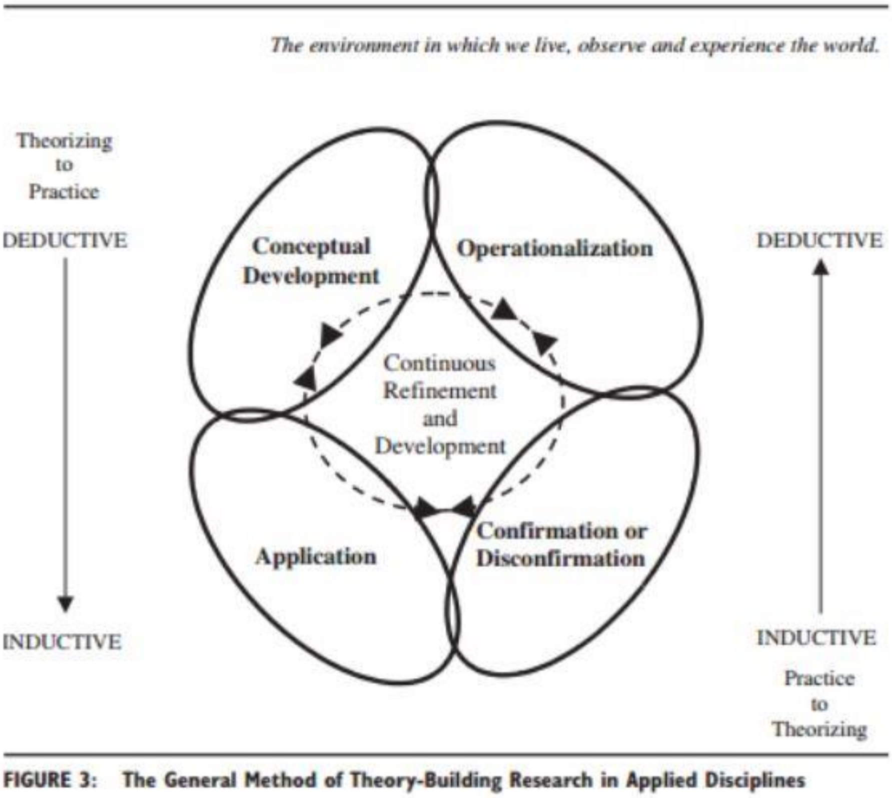

Figure 1: Applied research method according to Lynham (Lynham, 2016)

Therefore, this research chooses two different approaches – Deductive and inductive.

A deductive approach to research design comprises the development of a theory that is subject to testing. According to Saunders, it consists of deducing a hypothesis, expressing the hypothesis in operational terms, testing the hypothesis, examining the specific outcome, and if required modifying the theory based on the findings. With this approach, the method is to explain causal relationships between variables, and utilizing controls to allow the testing of a hypothesis. (Saunders, Thornhill, & Lewis, 2009) .

The deductive approach attaches great importance to the possibility of measuring quantifiable facts in our hypothesis and controls to allow the testing of the hypothesis so that the results can be replicated. This is the case, as all facts in this research are numeric and can be all evaluated in the available dataset (Appendix 5).

Additionally, to ensure scientific rigour, the deductive approach dictates that the researcher should be independent of what is being observed, which the case in this thesis is. The researcher should solely analyse collected data and combine it with other sources of data. Although this suggests a high level of objectivity, the selection of cases analysed constitutes possible bias as the question of what is and what is not a business case is very subjective and different people with different backgrounds may create a different dataset. Therefore, a clear definition of 'business case' is necessary for the correct replication of this research. This follows from the principle of reductionism, which holds that problems in their complexity are better understood when they are reduced to the simplest possible elements. (Saunders, Thornhill, & Lewis, 2009)

The final requirement of the deductive approach concerns generalisation. It is necessary to base our findings on a dataset of a sufficient size to support our findings. (Saunders, Thornhill, & Lewis, 2009)

Although the importance is mainly on the quantitative approach, this research only uses the quantitative findings to support the qualitative findings. With the research question of this thesis asking the how and what questions, it has however some characteristics of the inductive approach which builds the theory on observations. In this case, the observations provide information about factors influencing the Court, which are not present in the data used in the case. The researcher therefore uses the freedom given by this framework to form a hypothesis.

The inductive approach seeks to gain understanding of meanings humans attach to events and their context. It provides a more flexible structure to permit changes of research emphasis as the research progresses, as well as realisation that the researcher is part of the research process. However this flexible structure also means that one has to live with the fear that no useful pattern emerges. (Saunders, Thornhill, & Lewis, 2009)

By combining the methods mentioned above, one gets an approach which aligns perfectly not only with data collection but also with the philosophy of this thesis.

iii) Case study

In order to analyse all aspects of the research question, we need a case study, which provides for *"empirical investigation of a particular contemporary phenomenon within its real life context using multiple sources of evidence".* A case study is very suiting as this thesis seeks to answer how, what, and why questions, and acknowledges that the boundaries between the subject matter and the context within which it is being studied are not clearly evident. (Saunders, Thornhill, & Lewis, 2009). Especially important is the advantage of using triangulate multiple sources of data, which allow for verification of each source by the rest of the sources used.

There are many types of case studies, differentiating in the number of cases included in the research, and also in the unit of analysis. Due to the unique character of the example, this research follows the single, embedded case study as no other example can be compared to the present model and their environment is important for the analysis of the research question.

This research benefits therefore from the approach of a case study through the numerous sources it can use, the stress on uniqueness as well as the possibility to explore its multi-dimensional approach.

The case study is the prediction model created by Daniel Martin Katz, Michael J Bommarito II, and Josh Blackman. It is the perfect example for it is the first fully predictive and general model forecasting the decision outcomes of the US Supreme Court, based on empirical evidence, perfectly mirroring the factors involved in the decision making.

The prediction algorithm not only predicts the outcomes with a relatively high accuracy, it also shows the most important factors in the decision making, all based on a database publicly available. High accuracy and the robust character of the model by Katz et al. shows that the model could be potentially applied elsewhere.

iv) The data

Critical realism philosophy and the combination of the deductive approach indicate the use of both qualitative (non-numerical) as well as quantitative (numerical) data.

Not surprisingly, there is an increasing trend of using multiple data collection methods, which give space to creativity of the researcher. The approaches do not pose many limitations as to how the qualitative and quantitative data is to be analysed, most liberal approach being the mixed-model research, which

> *"[c]ombines quantitative and qualitative data collection techniques and analysis procedures as well as combining quantitative and qualitative approaches at other phases of the research such as research question generation."* (Saunders, Thornhill, & Lewis, 2009)

This thesis uses this liberal approach because it gives the author the possibility to exploit the complementary character of qualitative and quantitative character. Moreover, the research uses different sources of information which highly benefit from the possibility to analyse them qualitatively as well as quantitatively.

This research uses primary sources, such as the model, and its outcomes, which are quantitative. These sources are used to validate statements from literature and herewith the hypothesis, which is validated with a machine learning approach.

The research uses also secondary sources, such as the academic literature, which help framing the hypothesis and understanding the advantages of Data Driven Decision Making.

Last but not least, some information was also collected through correspondence with one of the authors of the algorithm predicting outcomes of the US Supreme Court, which enabled the author to run the code.

v) Manipulation with data

The dataset used in the prediction model by Katz et al. (Washington University Law, 2016) was unfortunately missing the possibility to filter required information for the analysis of this research. According to the article by Lee Epstein et al., many business cases are not in the 'economic activity' category and even if combined with other relevant categories, which should include all relevant business cases, 44% of business cases were left out. (Epstein, Landes, & Richard, 2016)

Therefore, the author decided to filter the database by Chief Justice, leaving only cases decided in the Roberts Court, which were then analysed on business relevance on a case by case basis. Any case, which would not have a direct impact on business would be removed from the dataset. The cases could not have been filtered just by selecting businesses as petitioners or respondents. In some cases, none of the parties involved was a business.[1]

Further, a few features were manually added to the filtered dataset from supremecourt.gov and justice.gov/osg/supreme-court-briefs.

vi) Cross sectional study

This research concentrates on a particular phenomenon today and is therefore referred to as cross sectional, because it seeks to answer the question how can Machine Learning help understanding the decision making of the US Supreme Court. The past as well as the future play an important role in the present and an essential part in the understanding of any phenomenon. However as opposed to longitudinal study, which studies change and development, this research is interested in the situation today.

[1] i.e. Pat Osborn v. Barry Haley, et al; docket: 05-593
Arnold M. Preston v. Alex E. Ferrer; docket: 06-1463
William G. Schwab v. Nadejda Reilly; docket: 08-538

3) Decision making: Connection between Law and Machine Learning

This chapter aims to review relevant literature, allowing an easy understanding of the theories and methods used.

The first section gives an overview of the Supreme Court, its structure, processes and biases, explaining the complexities of the system and the need for a model, allowing for a simple review and better understanding. The second section provides an insight into decision theory, decision making and human limitations in that domain and finally a solution for decision inconsistency suggested by experts – Data-Driven Decision Making. The last section of this chapter then focusses on Data Science, more specifically, Machine Learning as a field, the relevant methods as well as presentation of the case model.

a) The United States Supreme Court

The majority of countries have a court that rules on the constitutionality of the lower courts. It is the highest court in the hierarchy judicial review and therefore receives the highest respect, which is why it is not subject to further review. Depending on whether it is a common law or civil law jurisdiction, the court receives more or less power.

Civil law jurisdictions tend to have multiple supreme courts, each for different kind of geographic or subject area in order to avoid abuse of power. Additionally most of the European Civil law jurisdiction also have separate constitutional court, first developed in the Czechoslovak constitution in 1920 (Langášek, 2011). One of the civil law countries is Germany, with five different supreme courts, each dedicated to a different area.

In jurisdictions where the doctrine of stare decisis applies i.e. Common Law, the Supreme Courts get more power as stare decisis has the power to create precedents that are to be followed by lower courts. Consequently, their power exceeds the one of judicial review to partial legislative power.

The United States of America belong to the common law jurisdictions. As any federation, it has different levels of government - federal, state and local. The federal government has three branches – the executive (the President), the legislative (shared between by both of the houses of Congress) and the judicial (the Supreme Court). This separation of powers was established in the case of the United States in the Marbury v. Madison case (1803) in order to avoid misuse of power. At the time however, the power of judgment was underestimated. (Hawley, 2012)

i) The composition of the Supreme Court

The composition of the Court, is fixed by rules and conventions. The most important requirements are listed below:

- **Plural membership** – The Court has to be composed of at least 2, probably 3 members. This number was changed a few times by the Congress, the last time in the Judiciary Act of 1869 to 9 members (eight associate justices and a Chief Justice). It is emphasized to have a great variety of characters to represent the whole country.
- **Life service in Office** – The members are subject to good behaviour. They can therefore not be removed from the office except by impeachment. This requirement is there to ensure independence of the Justices. It is the highest judicial review in the hierarchy, where people expect impartiality and justices should therefore not fear for their job every time they make an unfavourable decision, which is why the life service office is in place.
- **Proficiency in Law & Practical affairs** –The job is quite demanding when it comes to volume of work and time pressure, it is the highest instance of the judicial branch and deals primarily with administrative law and the most complex cases subject to controversy, which explains why the majority of justices appointed to the bench had a law degree. It is not a requirement, but a proven discernment in matters of complexity, steadiness in matters of difficulty, and coolness in matters of controversy is a must.
- **Constituency** – The Court does not have electoral constituency as it is not directly nor indirectly elective and the direct involvement with the political process is strictly proscribed. The reason for that is that the Justices have to be driven solely by their expertise and their belief, not their political drive. (Hazard, 1978)

According to the United States Constitution, justices are to be nominated by the President of the United States of America (hereafter POTUS), the Senate than has to approve of the nominee and if the Senate consents, the justice is then appointed to the Supreme Court bench. In exceptional cases, when the Senate is in recess, the POTUS may make temporary appointments.

ii) Process

Article III of the United States Constitution 1789 gives the Supreme Court the capacity to establish judicial review, herewith, to declare unconstitutional the actions of Congress as well as the legal provisions of state and local governments.

All filed cases fulfilling the requirements given by the Constitution are read through by the clerks of the Supreme Court and only a few are selected for further review, usually mainly because of disagreement between lower courts. However each justice has different ideas about what case needs to be reviewed, and although there are many legitimate reasons for review, they can only take on about 80 cases per term, which is why the justices vote on each case and only take it if a minimum of 4 justices agree. The whole process is regulated by *'Rules of the Supreme Court of the United states PART III. Jurisdiction on Writ Of Certiorari'*

The number of cases brought to Supreme Court exceeds the capacity of the Supreme Court, and only a small percentage can be granted certiorari. The official website of the US Supreme Court states that

> *"The Court receives approximately 7,000-8,000 petitions for a writ of certiorari each Term. The Court grants and hears oral argument in about 80 cases."* (The US Supreme Court, 2016)

This is only about 1 to 1, 14%. About 99% of all cases submitted to the Court are therefore not addressed. According to Amanda C. Bryan, there are a few predictors, which increase the likelihood of a grant. Under these predictors counts the existence of strong conflict and the number of amicus briefs. The grant recommendation from the Solicitor General or the lower court exercising the power of judicial review.

According to the article, *"When the Solicitor General Recommends a grant, the probability of a grant recommendation more than doubles."* (Bryan, 2016)

Once that petition is granted, each of the parties to the litigation (petitioner and respondent) submits written materials presenting relevant issues. More and more common are amicus briefs submitted at this stage.

According to an article posted by Adam Feldman:

> *"Few cases now come through the Court's merits pipeline without amicus briefs filed on the merits, and a growing number of granted (and denied) cases have amicus briefs filed at the cert stage as well."* (Feldman, 2016)

Moreover, an article by Brandon D. Harper states that in the 2011-2012 term, there were more than 300 amicus briefs filed, while only 83 cases were heard in that term. (Harper, 2016)

When all materials have been presented, the case is scheduled for a hearing, where each of the parties provides oral arguments before the Court. Based upon the weight of these arguments and other factors, each participating justice ultimately cast his or her vote whether to affirm or reverse the decision of the lower court.

In order for the review to have an effective outcome, there needs to be a majority as cases are decided by majority vote of the Justices. It is important to note that the fact that the case was argued in a specific term does not mean that the opinion will also be delivered by the end of that term. It does also not mean that the case will be decided. In exceptional cases, after the hearing, justices come to the conclusion that the certiorari should not have been granted in the first place.

There are many different outcomes of a case, mostly because it is a Supreme Court, which handles the most unique cases. They however all end at the lower court, where the case is reheard and further dealt with according to opinion delivered by the justices. The extent to which the case has to be reargued is specified in the art of opinion.

Affirms, which affirms the lower court's verdict. This decision is made if the decision is not clearly erroneous.

Vacate, which is delivered in cases, where the court did not correctly apply the law to the facts of the case or if the Court responds to a different question, not mentioned in the previous decision. In this case, it can happen that the Court reverse and remands the lower court's decision. In both cases the Court sends the case back to the lower court and sends explanations as to how the case should be dealt with.

Reverse, where the Court reverses the judgement made by the lower court in total. In that case, the Court disagrees with the lower court's entire decision.

It can happen that the court did not completely disagree with the lower court's decision. It only disagrees with a part of that decision, this means that only a part of that decision will be reversed. In some cases, it can also be **remanded with instructions**. The reason for reversing in part lies in the costs, instead of rehearing the case all over again, it can rehear a part of the case and decide only on that particular part.

Another option is that the certiorari was originally granted but the justices then come to the conclusion that it was not supposed to be and therefore **annul the certiorari.**

Before any case arrives at the Supreme Court, it often has to go through a very long process of endless appeals.

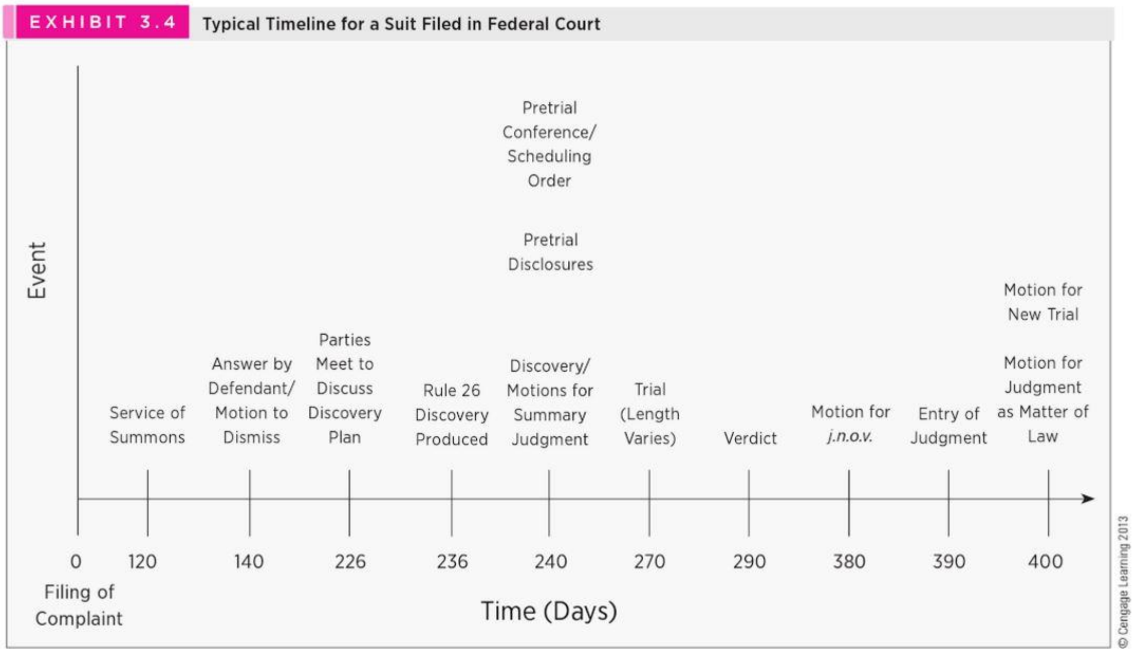

Figure 2: Typical Timeline for a Suit Filed in Federal Court (Bagley, 2012)

The process is not just time-consuming and long, it is also very expensive due to the amount of administrative tasks it includes, not to mention the attorney fees, therefore many cases worth ending in the Supreme Court never make their way up due to the time constraints and the expensive character of such matter. It is important to keep that in mind when analysing the Court's behaviour. Many cases never make it to the Court and the case is rather settled and there is currently no database including these cases.

iii) Different factors influencing the outcome

The outcome is influenced by different factors, such as the facts, the constitutional interpretation of the Justice (originalism, pragmatism, textualism, constructionism, structuralism, etc.) and their personal beliefs.

As mentioned in the Introduction, the outcome is bias by the justices' background and upbringing. When looking at the details about each justice currently sitting at the bench, one remarks one thing – they are all part of an elite, because sound education is a prerequisite for judicial service. Even though the Court should reflect the opinion of the whole country, the Court clearly represents the wishes of the elite and some cases coming to Court may suffer from the lack of diversity. They all have a background, most of them in Government positions which contributes to their bias for or against businesses.

An article by Richard A. Posner stresses the *"importance of interest groups pursuing private goals rather than public interest, in shaping legislation".* (Posner, 1986)

This can be done particularly with the help of amicus briefs, which are submitted to the Court. An Amicus curiae brief is defined by the Rules of the Supreme Court of the United States, Part VII, Rule 37.[2]

"*It is believed that they reduce information problems at the Court by helping the Justices anticipate the impact of their opinions."* This statement from 1992 was tested and was found largely inaccurate, because amicus briefs often contribute unique arguments and they also commonly reiterate their party's brief, which is why amicus briefs deserve much attention (Spriggs & Wahlbeck, 2016).

Since Justice Roberts, who filed many amicus briefs before joining the Court, took the place of Chief Justice, amicus briefs deserve even more attention than ever before (Harper, 2016).

The person or organisations filing the briefs also matters as Justice Ruth Ginsburg reported in her interview:

> *"[T]he experience of the attorney (particularly experience before the Supreme Court) would be a likely barometer of the quality of the arguments set forth in the brief."*

Where

---

[2] *'1. An amicus curiae brief that brings to the attention of the Court relevant matter not already brought to its attention by the parties may be of considerable help to the Court. An amicus curiae brief that does not serve this purpose burdens the Court, and its filing is not favored. An amicus curiae brief may be filed only by an attorney admitted to practice before this Court as provided in Rule 5.'*

'4. No motion for leave to file an amicus curiae brief is necessary if the brief is presented on behalf of the United States by the Solicitor General; on behalf of any agency of the United States allowed by law to appear before this Court when submitted by the agency's authorized legal representative; on behalf of a State, Commonwealth, Territory, or Possession when submitted by its Attorney General; or on behalf of a city, county, town, or similar entity when submitted by its authorized law officer.'[2]

“[H]er clerks often divide the amicus briefs into three piles: those that should be skipped entirely, those that should be skimmed, and those that should be read in full.” (Epstein, Landes, & Richard, 2016)

Amicus briefs filed by an attorney with experience thus have an impact on the decision of the Court whereas amicus briefs filed by an attorney without experience are more likely to be skipped than not.

State Attorneys or Solicitor General have a clear advantage as they have more experience and herewith chance to have their arguments heard. The current Court’s inclination to favour amicus briefs filed by the Government can also be explained by their past where the majority of the sitting justice had experience in the Government functions.

Justices with Significant Government Experience	
Roberts	Deputy/Acting Solicitor General
Scalia	Solicitor General Prospect
Thomas	Assistant Missouri Attorney General
Breyer	Special Assistant United States Attorney General
Alito	Assistant to the Solicitor General
Sotomayor	Assistant District Attorney
Kagan	Solicitor General

Justices with Significant Private Sector Experience	
Kennedy	Law Professor
Ginsburg	ACLU General Counsel

Figure 3: The Effectiveness of State-Filed Amicus Briefs at the United States Supreme Court (Harper, 2016)

Moreover, according to an article by Lee Epstein et al., in addition to favouring briefs and opinions by the Solicitor General, these opinions tend to support business. (Epstein, Landes, & Richard, 2016) Finally, as a consequence of the current Court being pro-business, the increasing number of amicus briefs filed by the Chamber of Commerce, assuring evolving experience of their attorneys, briefs filed by the Chamber of Commerce are most likely also carefully read.

The Supreme Court is a complex system, which is why the factors mentioned above give the scholars an idea, but are not in any way exhaustive. It is almost impossible for human beings to consider all the different factors and attach to each the correct weight that would reflect its importance

iv) Bias on the Supreme Court

The US Supreme Court, as any other institution, is subject to bias, which have to be analysed and addressed in order to prevent undesired outcomes. The main potential biases come from the decision, what cases will be heard. Only the cases that actually receive petition for certiorari have the chance to make an impact. There is only a little information about what cases did not make it and the reason why, still these cases represent about 99% of the total cases filed. This task is usually handled by the Cert. pool – a shared legal labour pool that streamlines the work of reviewing incoming cases, leaving most of the decision in hands of justices' clerks - to be exact, with only one of them, as work is shared among all of the clerks, excluding clerks of Justice Alito, who do not participate and review all cases. (Liptak, 2008)

Biased recommendation from the Cert. Pool memo could therefore bias the entire Supreme Court, preventing it from taking cases, especially since the majority of the justices participates in the Cert. Pool and receives the same memo, and grant of certiorari only requires 4 votes. (Bryan, 2016)

Another bias included in this institution is the extent to which the case will be decided. When submitting cases, a clear question(s), that need to be argued before the Court has (have) to be formulated. The Supreme Court has the freedom to set the definitiveness as well as fullness of its decision, which can end up with a case being decided in a way that still leaves many important questions unanswered. (Toobin, 2008)

One of the most discussed and probably the most obvious biases is the composition of the Supreme Court and the background of each of the justices, which has a great impact on any decision as no decision can be made without taking into account ones background.

Probably the most complex bias is the one coming from the Game Theory. Each of the justices has an idea about the correct outcome of a case and they are all more or less equal players, which means that no one has control over the outcome, if the Chief Justice and his power to organise the program is not taken into account. The decision then stays with the decision of each individual justice. Based on each justice's priorities, they can play the game theory in their favour, by going with a decision that does not exactly correspond their belief and writing the opinion in a way which softens the undesired consequences of the outcome. (Toobin, 2008)

v) The dangers of the Supreme Court

> *"Few other courts in the world have the same authority of constitutional interpretation and none have exercised it for long or with as much experience."* (Hughes, 2016)

Theoretically, in case of abuse, the executive could refuse to enforce judgements he did not like, however due to the equality given into each of the branches, this scenario is unlikely. According to former law clerk of the sitting Chief Justice Roberts, Hawley,

> *"if the judiciary can count on regular enforcement of its decisions by the president and respect for them by Congress, its power of judgment becomes just as weighty in the constitutional scheme as the executive's power of the sword or the legislature's control of the purse — quite contrary to Hamilton's claims."* (Hawley, 2012)

Suggesting that the power of the judicial branch may be greater than the power of the other two branches, because once an opinion on a constitutional issue is announced (Marbury v Madison), it cannot be altered by any vote or procedure. The fact that justices are appointed to lifetime service and the only way to remove them from the bench is by impeachment gives them *"final say on every constitutional matter it adjudicates."*

This is especially dangerous, because if the Supreme Court deploys its judgement authority entirely within its legal limits, it becomes the constitution's primary interpreter as well as enforcer, because the American judges are independent in the full sense of the word independent as they are not subject to review and their decision cannot be altered. (Hawley, 2012)

This would not be an issue, if the Court did not decide on nearly every political question. For the first century of its existence, the Court stayed mostly clear of constitutional questions, with the time however it got very comfortable interpreting rights and in the middle of the 20^{th} century invalidating federal and state laws. Moreover, according to Hawley, they no longer use the text of the Constitution to make their decisions, instead they favour their own moral predilections and policy views. (Hawley, 2012)

According to Robert A. Dahl, The Supreme Court is also seen as a national policy-maker.

> *"What is crucial is the extent to which a court can and does make political decisions by going outside established 'legal' criteria found in the precedent, statute and constitution."* (Dahl, 2016)

Amy Howe explains this problem in her article '*Interpreting the Supreme Court: Finding Meaning in the justices Personal experiences'.*

> *"In the absence of clear right or wrong answer, the Justices often have to operate in a grey area. (…) The key is to look at a Justice entire experience collectively, because that is what the Justice will rely on to make decisions and that is what will inform how she sees tough questions that the Court decides"* (Howe, 2015)

Nevertheless is it a big issue, as the Court does not rely solely on the Constitution. The decision is influenced by other factors as well, which are not defined and it is up to the justices to attach the right weight to each of them.

vi) Roberts Court

Roberts Court is the Court under the leadership of Chief Justice Roberts, the 17^{th} Chief Justice, it started in 2005. It is considered to be one of the more conservative Courts due to the political preferences of its members. According to an article by Paul Barrett, published in December 2015, the Court sides predictably with corporations in their clashes with consumers (Barrett, 2015). Noam Scheiber goes even further by stating:

> *"[S]ome argue that the Supreme Court under Chief Justice Roberts has become perhaps the most business-friendly court in recent history."* (Scheiber, 2016)

According to a 2013 study by Lee Epstein, where justices were ranked according to their rulings in cases involving business, *"six out of the 10 most business-friendly justices since 1946 sat on Supreme Court at the time of Justice Scalia's death."* (Scheiber, 2016)

Conservative appointees tend to side with businesses while democratic appointees sides with the other party involved, or so one appears to think. The article by Scheiber also stated:

> *"Democratic appointees to the court have reliably dissented on the most sweeping pro-business decisions in areas like class actions and arbitration. But they have frequently taken pro-business positions in other rulings."*

The simple indication that a justice has been appointed by a democratic president does not mean that it will automatically side with dissent in pro-business decisions. A prediction of a case can therefore be more complicated than just analysing the background of the justices. The president can appoint a candidate who does not share his political or ideological views, but is the best choice to advance his goals *"They search for nominees, who are political allies, not political pawns or prodigious legal minds."* (Epstein, Martin, Quinn, & Segal, 2007)

The Court consists currently of eight justices (see Appendix 1) with one vacant position due to the sudden death of Justice Scalia. The Justices are appointed by different presidents, whose interests and personal values they usually represent. The appointed Justices then often show loyalty towards the president who appointed them. (Epstein & Posner, epstein.wustl.edu, 2016)

Note that the majority of the Court is conservative, belongs to the Roman Catholic Church and was educated at the top American universities, making undoubtedly part of the elite.

b) Decision-making

"Decision' implies the end of deliberation and the beginning of action." William Starbuck (Buchanan & O'Connell, 2006)

"The question, who makes what decisions and how have shaped the world's systems of Government, justice and social order" The answer includes different intellectual disciplines, such as mathematics, sociology, psychology, economics and political science for example. The business oriented answer is however more oriented towards risk: How can managers achieve better outcomes? (Buchanan & O'Connell, 2006)

According to March and Simon (1958), managing organizations and decision making are virtually synonymous, which explains why we keep hearing questions about decision making in connection with organisation' management. (Nutt & Wilson, 2010)

Today's understanding of managerial decision making has its origins in decision theory and information theory, defined by theorists such as James March, Herbert Simon, and Henry Mintzberg. The decision making is therefore concerned with identifying the best solution, with taking decisions under uncertainty and decision-making by a group of people.

According to Simon, human beings would make economically rational decisions if only they could gather enough information, which aligns with Erik Brynjolfsson, who says that more precise and accurate information should facilitate greater use of information in decision making and therefore lead to higher firm performance. (Brynjolfsson, Hitt, & Kim, 2016)

Mintzberg et al. (1976) state, that

> "*[S]trategic decisions are seen as large, expensive and precedent setting producing ambiguity about how to find a solution and uncertainty in the solution's outcomes. Once implemented, a strategic decision stipulates premises that guide operational decisions that follow.*"

A person making a decision then takes a position of a historian, digging deep into the past to see whether a similar issue has occurred and how it has been solved, looking how it could be solved and calculating the chances that the preferred option will make profit and will not jeopardise the whole company.

According to an article by Salesforce '*AI for CRM: A field Guide to everything you need to know*'

> *"Half of all business decisions are made with incomplete information, which disconnects the business from the product and therefore from the customer."* (Salesforce, 2016)

Decision making is setting values, objectives and goals, which are the key to any business. Choosing the wrong values, objectives and goals can ruin the whole business, therefore better strategies need to be developed to reduce uncertainty, and improve the accuracy of the predictions based on information available at the time of the decision.

i) Humans and decision-making

The majority of decisions are currently being made by humans. This is an important problem, because humans are unreliable decision makers; their judgements are strongly influenced by irrelevant factors such as their current mood, the time since their last meal and the weather. (Kahneman, Rosenfield, Gaudhi, & Blaser, 2016) They do not formulate utility functions, the pursuit of which they are expected to conduct with relentlessness and consistency. Moreover, the fallacies of human's fast thinking is that the fast survival process only looks at the information at hand and does not actively look at the additional information. Furthermore, it does not weigh the different components of the information according to their objective importance, but how vividly they appear in human's mind.

According to the article '*Noise: How to overcome the high, hidden, cost of Inconsistent decision making?*', professionals often contradict their own prior judgements when given the same data on different occasions, furthermore, if two different professionals are to judge the same data, the differences are even more likely to diverge. This shows great inconsistency in the way humans make decisions. (Kahneman, Rosenfield, Gaudhi, & Blaser, 2016)

This difference, called noise, is said to be the reason why companies keep losing money, without even realising it. Even if the noise was to be reduced only by a few percentage points, the value of such an outcome would be measured in tens of millions. (Kahneman, Rosenfield, Gaudhi, & Blaser, 2016)

An introduction of Machine Learning in the process would therefore reduce this bias of readily available information and make humans focus only on the relevant information. (Kahneman, Rosenfield, Gaudhi, & Blaser, 2016) It would also check against the tendency to rely on subjective weights. By getting a list of attributes listed by weights, one can resist the seduction of constructing an attractive narrative, which is so typical in human decision making.

(a) Group decision making

Group decision making is more respected because it includes people using all their wisdom to make acceptable and fair decisions for everyone, which requires certain respect for everyone in the group. The *'field theory'* by Kurt Lewin says that actions taken by the group are determined by social context and although each member has a different opinion, they all work together to achieve the same goal. According to Peter Drucker, the most important decision is not made by the group itself, it is made by the management, which chooses the group. A dynamic is therefore essential in the decision making of a group and can influence its outcome. (Buchanan & O'Connell, 2006)

ii) Data Driven Decision Making

Through our education, we are told to trust our guts and our instinct and we are led to believe that with experience we can make accurate assessments. We are told to trust our judgment – which is wrong! In fact, Andrew Mcafee describes this as the biggest challenge of big data, and the most harmful misconception in the world, and asks a naïve question: How do we convince people not to trust their own judgment? (McAfee, Big Data's Biggest Challenge? Convincing People NOT to Trust Their Judgment, 2003)

Paul Meehl, who started comparing human experts to algorithms almost 60 years ago concluded at the end of his career:

> *"There is no controversy in social science which shows such a large body of qualitatively diverse studies coming out so uniformly in the same direction as this one. When you are pushing over 100 investigations, predicting everything from the outcome of football games to the diagnostic of liver*

> *disease, and when you can hardly come up with a half dozen studies showing even a weak tendency in favour of the clinician, it is time to draw a practical conclusions".* (McAfee, Big Data's Biggest Challenge? Convincing People NOT to Trust Their Judgment, 2003)

There is no question, whether it would give us better results. The theory and the evidence speaks for algorithmic solution.

According to a study conducted by William M. Grove et al., which compared the clinical and mechanical data combination technique, used to making judgements and decisions, mechanical prediction techniques were about 10% more accurate than clinical predictions. Mechanical prediction then substantially outperformed clinical prediction in 33-47% of all studies examined. There were only a few cases, where clinical cases outperformed, and in these 8 cases, only one consistent pattern emerged. In the 7 out of 8 cases, the clinicians received more information than what was available to the mechanical prediction, while the superiority for mechanical-prediction technique was consistent independently of the judgment task, type of judges, experience or the types of data being analysed. (Grove, Zald, Lebow, Snitz, & Nelson, 2016)

Although Machine Learning has been here for decades, it never achieved its full potential. This is changing thanks to Moore's Law, which states that processor speed increases and Kryder's Law, stating that the price for data storage declines. Machine Learning finally starts to be taking place in every part of our lives, even the ones often attributed to experts. In 2011, a team from Stanford used a set of 248 breast cancer samples from women in the Netherlands to train the program to identify severity of the disease. The team manually trained the model to make difference between different types of tissues. They also let the program draw its own conclusions, which led it to identifying 6642 important features for the images. These features included standard elements, such as morphometric descriptors of image objects and higher-level contextual, relational and global image features. Some of the features were completely irrelevant, some of them were identified by pathologists as standard identifying features, but what was surprising was that it identified features, which are highly relevant, but until that moment completely unknown to pathologists. (Beck, Sangoi, Leung, Marinelli, & Nielsen, 2016) (McAfee, Andrew McAfee, 2016)

This shows, that Machine Learning outperforms even medicine experts, in a task, where precision, accuracy as well as consistency are the most important factors. Factors, where human beings do not perform so well.

What the theory and evidence is telling us is that we are all replaceable. But according to an article by Andrew McAfee, the situation is not as black as we might think. Yes, human beings have significant limitations. They do not perform well in messy, complex, real-world environments, because these areas do not allow for immediate feedback, or practice, needed for the development of intuitive experience. This is where according to Kahneman, algorithms perform much better, thanks to their ability to find outliers and to maintain a modest level of accuracy due to their ability to decide consistently. (McAfee, When Human Judgment Works Well, and When Iit Doesn't, 2014)

Humans also tend to ignore base rates, assign non optimal weights to cues, and fail to take into account regression toward the mean and also to properly assess covariation. They believe in the law of small numbers and over-weight vivid data. (Grove, Zald, Lebow, Snitz, & Nelson, 2016)

Even when using analytics to aid our judgment, we usually do worse than the model itself, because we second-guess the judgment made by the model. We do better than without the model, but still, the model itself would perform better (McAfee, Big Data's Biggest Challenge? Convincing People NOT to Trust Their Judgment, 2003), showing clear bias against models and lack of trust.

An interesting observations is that the outcome improves in quality as well as performance, when expert's subjective opinions are quantified and added to an algorithm. With a growing number of data available, the importance of human judgement should decrease. This means in no way that humans should be eliminated from the decision process. (McAfee & Brynjolfsson, 2012) While an increasing number of operations is happening automatically, human being still play a key role in the process of decision making. No matter how much we advance, human beings possess capabilities still unavailable to machines, which are the capability to think outside the immediate scope of a task or to deal with ambiguity. (Reeves & Ueda, 2016)

While machines excel at executing well defined tasks, human beings' strength is the creativity, which should be used to frame central questions around applications of Machine Learning, providing more transparency into the black box. Another strength of human judgment is the ability to challenge our thinking, making us continuously reframe our understanding of a problem. To make data driven decisions useful and valuable, these characteristics have to be implemented in the new models in order for them to be embraced and leveraged by organizations. (Reeves & Ueda, 2016)

The approval of organisations is still lacking, but it is soon to change because as W. Edwards Deming used to say: *"You can't manage what you don't measure."* Data-driven decisions lead to improved decision making and performance. Not only do they lead to better performance, according to a study conducted by MIT Center for Digital Business, companies that characterized themselves as data-driven performed better on objective measures of financial and operational results. *Companies in the top third in the use of data driven decision making were on average 5% more productive and 6% more profitable than their competitors.* (Reeves & Ueda, 2016)

Finally, there is growing evidence that data driven decision making has a positive impact on firm's performance. *"A research proving the theory, concludes that DDD capabilities can be modeled as intangible assets which are valued by investors and which increase output and profitability."* (Brynjolfsson, Hitt, & Kim, 2016)

c) Data Science

At a high level, data science is a set of fundamental principles that support and guide the principled extraction of information and knowledge from data. It is a young field, concentrated around extracting information from data. Information in this context

> "*[i]s a quantity that reduces uncertainty about something. The better the information, the more my uncertainty is reduced"* (Provost & Fawcett, 2013).

Through Statistical and Machine Learning methods, **it aims to discover unknown patterns, predict future outcomes or explain past events.** It seeks to gain insight through the application of approved "*scientific methods, that is, empirical and measurable evidence subject to testable explanations and predictions.*" (Wladawsky-Berger, 2016) The field is large and includes domains such as Machine Learning, Data Mining, Artificial Intelligence and Big Data. It also includes older domains such as statistics and analytics, which is why most of the models used in Machine Learning come from that field. The best quality of data science is that it can be applied to various different domains due to our ability to collect data almost about any aspect of our environment.

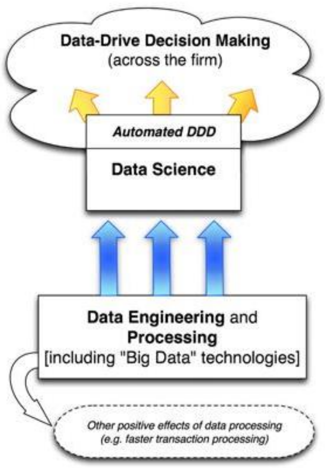

Figure 4: Data science in the context of closely related processes in the organization. (Provost & Fawcett, 2013)

Through the collection of data and leveraging informative insights, its ultimate goal is improving decision making, as this generally is of paramount interest to business.

i) Machine Learning

This research concentrates on Machine Learning. Machine Learning as a field of study arose as a subfield of Artificial Intelligence (AI), which was concerned with methods for improving the knowledge or performance of an intelligent agent over time, in response to the agent's experience in the world (Provost & Fawcett, 2013). Generally, Machine Learning consists of two parts: Supervised and Unsupervised Learning.

Unsupervised Learning, also called descriptive is concerned with finding patterns in a dataset and herewith bringing structure into an unlabelled dataset. Datasets used in unsupervised learning do not have any target variable and are therefore segmented into clusters based on similarities in the data. Under unsupervised learning count methods such as clustering, neural networks or principal component analysis for example.

Supervised learning on the other hand, is an area of Machine Learning, where the analyst is trying to build a model behind a dataset. The analyst is usually trying to make predictions about the future data based on already collected data. Such predictive model is then a formula for estimating the unknown value – the target variable.

One of the main characteristics of supervised learning is the target variable, which is already defined in the dataset. To the most common methods count Support Vector Machine, Classification Trees, K-Nearest Neighbours and many other methods.

Decision functions, which result from the training process, can transfer variable values into predicted labels. The prediction performance of these functions can be done by comparing the predicted values with the target variable. With an appropriate design/method, these functions can be applied to different problems. (Chi, 2009)

Below are the most common methods:

(1) Decision Trees

Decision Trees, first introduced in 1986 by John Ross Quinlan, are used in supervised learning. They are simple and useful when the main objective is correct understanding of the model. They are however not very competitive with the other supervised learning approaches, which is why they are usually used in combination with other methods, which improve model's accuracy.

Tree-based models can be used for regression (attempting to estimate each individual's numerical value) and also for classification problems, which attempt to estimate or to predict the class probability of each individual by studying the relationship between the dependent and the independent variable. The difference between regression and classification is in the nature of variables, whether they are quantitative or qualitative and whether we search to answer a '*what*' or a '*how much*' question.

Regression problems concern quantitative variables whereas classification problems concern qualitative also called categorical variables. Therefore, whenever the question is how much, regression together with least squares statistical learning method is used whereas when the question is what, a logistic regression can be helpful to solve classification problems.

Decision trees are segmenting the total number of observations into subgroups based on the observations similarities and their dissimilarities, herewith growing a structure resembling to a tree – induction task.

The main part of induction plays the training set of objects whose class is known. The induction task then seeks to develop a classification rule that can determine the class of any object from its values of the attributes. The classification is then expressed as a decision tree as in the figure below. (Quinlan, 1986)

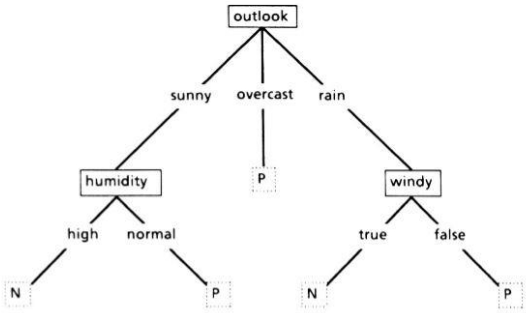

Figure 2. **A simple decision tree**

Figure 5: Example by Quinlan, 1986

Leaves of a decision tree are class names, while the nodes represent attribute-based tests, with each branch representing each possible outcome. Classification then starts with the root node and depending on the test evaluation, one takes the branch which corresponds to the appropriate outcome. This process goes on until a leaf node is encountered, which assigns the class of the observation. (Quinlan, 1986) This can lead to overfitting as with each node, the number of observations evaluated is reduced, ending with groups consisting of just one observation. This problem is resolved through pre- or post-pruning. Pre-pruning stops the growing process of a tree before all observations are perfectly classified, whereas post-pruning allows for perfect classification but once classified, the tree is pruned to improve the classification error on the test sample.

This method is very clear and easy to understand, however it is also susceptible to high variance and overfitting. To improve the accuracy of such models involves multiple trees, which are combined to yield a single consensus. This often results in remarkable improvements in prediction accuracy. Unfortunately there is a price to be paid for the improved accuracy - loss in interpretation. (Witten, Hastie, & Tibshirani, 2013)

(2) Random Forests

Random Forests (RF) is an ensemble method[3] for supervised and sometimes for unsupervised learning. It was first proposed in 2001 by Breiman, who defines it as:

[3] Ensemble method is a method, using multiple machine learning models to complement different parts of the data input and space and herewith complements each other's deficiencies.

"[A] classifier consisting of a collection of tree-structured classifiers {h(x,Θk), k=1, ...} where the {Θk} are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x ." (Breiman, 2001)

According to Ho, decision trees have many advantages, but the overfitting character causes a major issue as neither pre- nor post-pruning allows the trees to grow into arbitrary complexity while increasing both the accuracy of the training set and the accuracy of the testing set. (Ho, 1995) RF on the other hand, not only put multiple decision trees together. Each tree is grown with random subset of features. *"In building a random forest, at each split in the tree, the algorithm is not even allowed to consider a majority of the available predictors"* (Witten, Hastie, & Tibshirani, 2013)

The reason for that is that often, there are very strong predictors, which would otherwise be in the top split, causing the trees to look quite similar to each other and the correlation between each of the trees strong. As the variance and generalization error depend on the correlation, the goodness of the model would be affected.

One of the advantages of RF is that they are not as sensitive to overfitting as decision trees, because as more trees are added, they only produce a limiting value of the generalization error.

(3) Extremely randomized trees

This method was introduced in 2006 by Pierre Geurts, Damien Ernst and Louis Wehenkel. Although the RF and other tree based algorithm provide a randomization, they still result in a very high variance of decision and regression tree spits, which is why higher randomization was pursued with the aim to reduce variance coming from tree-based methods.

Extremely randomized trees is an ensemble method as one can deduct from the name. It consists of strongly randomizing both attribute and cut-point choice while splitting a tree node. The attributes are therefore selected randomly without considering their relationship to the output. The probability of selecting an irrelevant variable is therefore as same as the probability of selecting a relevant variable. This negative effect is then cancelled by averaging. The irrelevant variables are only there to reduce the size of the sample. (Geurts, Ernst, & Wehenkel, 2016)

The strengths of this method is accuracy as well as computational efficiency. (Geurts, Ernst, & Wehenkel, 2016)

(4) Support Vector Machines

Support Vector Machines (SVM) is a classification method developed in the 1990s by Boser, Guyon & Vapnik, which can also be used for regression. It is rooted in the Statistical Learning Theory developed by Vladimir Vapnik and introduced in the 1960s. The method is very robust in respect to outliers as it maximizes the margin between different (usually 2) classes with the help of two hyperplanes, placed at the edge of each class, H1 and H2 (also called support vectors), creating a clear boundary. (Boser, Guyon, & Vapnik, 1992) The resulting classification function then depends only on so-called supporting patterns. The whole idea is to find a function that has the least significant deviation from the actual targets, creating a large field, where any misclassification is seen as an error. According to Provost & Fawcett, it can of course happen that data is not linearly separable and the best model is then a balance between a fat margin and a low total error penalty. It is important to state that the penalty is proportional to the distance from the decision boundary (the support vectors).The accuracy of this model is therefore very good and the model received through this method is robust. (Provost & Fawcett, 2013)

A great property of SVM is its use of kernels. Kernels are very important when dealing with high dimensions and non-linearity.

Often, data in question is not linearly separable and there is no way to separate two classes, as in figure 5.

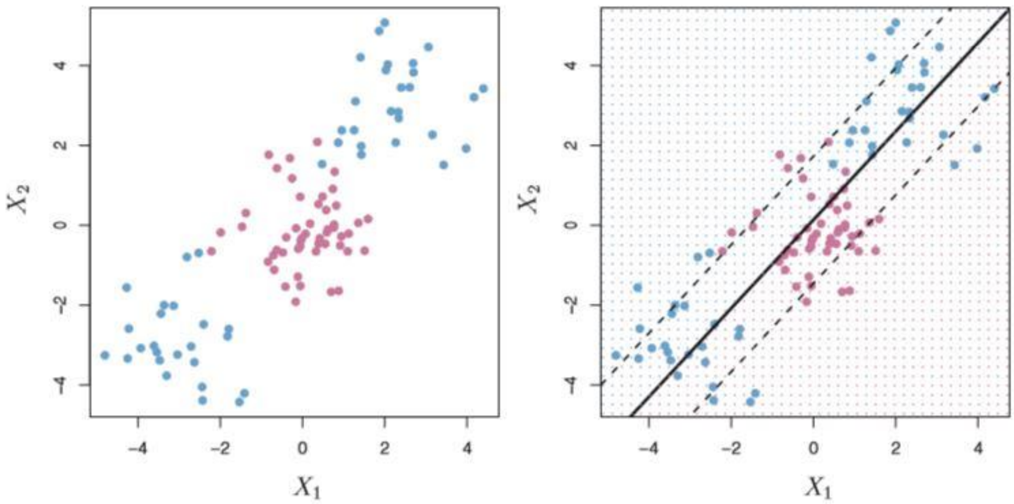

Figure 6: "Left: The observations fall into two classes, with a non-linear boundary between them. Right: The support vector classifier seeks a linear boundary, and consequently performs very poorly." (Witten, Hastie, & Tibshirani, 2013)

To achieve clear separation, the data is transposed or mapped into a higher-dimensional space, where the two classes can be easily separated with a linear function. This transposition is possible thanks to kernels, which are to be thought of as a function that quantifies the similarity of two or more observations. Popular non-linear kernels for this use are Polynomial kernels or Radial Basis Function (RBF) kernels.

Polynomial kernels are mapped through replacement of each instance with a weight and RBF kernels use infinite feature space. (Witten, Hastie, & Tibshirani, 2013)

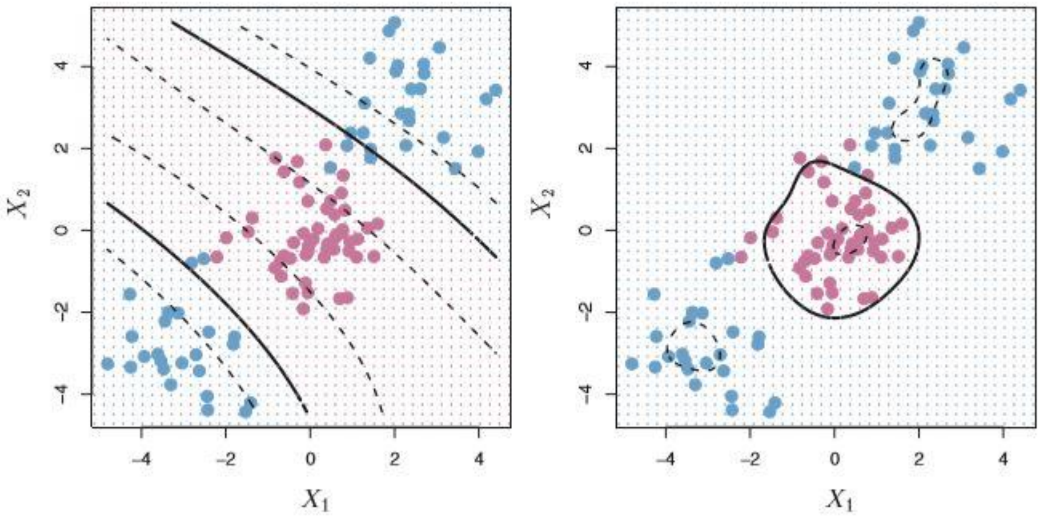

Figure 7: Left: An SVM with a polynomial kernel of degree applied to the non-linear data from Right: An SVM with a radial kernel is applied. In this example, either kernel is capable of capturing the decision boundary. (Witten, Hastie, & Tibshirani, 2013)

Besides its advantage of enabling linear separation, they are also computationally cheaper because kernels only need to compute K for all distinct pairs. No working in the enlarged feature space is required. (Witten, Hastie, & Tibshirani, 2013)

A potential bias from the use of kernels is the risk of overfitting due to better adaptation of the two hyperplanes. They have however many advantages, which is why they are used in SVMs but also in other machine learning methods.

(5) Validation

In order for the models accuracy to be measured, the original dataset is split into two sets. The training set and the testing set. The training set is used to build the model, whereas the testing set is set aside during the building process and used only at the end to monitor misclassification and other metrics. (Witten, Hastie, & Tibshirani, 2013) If the whole dataset was used for training, the accuracy of the model would then achieve perfect predictability, which would be misleading as it would not be tested on unseen observations.

A better validation method achieving low variance and low bias is k-fold cross-validation which uses k−1 partitions of the data for training and one fold for testing and repeats the process for k times.

There are also other methods, which do not require the creation of two or more different samples, such as bagging, used in the RF. In bagging, trees are repeatedly fit to bootstrapped subsets of observations, consequently each bagged tree makes use of around 2/3 of the observations and the rest is then referred to as an out-of-bag observation (OOB). The OOB error of the random forest is then calculated on the part of the dataset that was not used in the building of the model and represents a valid misclassification estimate. (Witten, Hastie, & Tibshirani, 2013)

These techniques are not only important for the validation of the model, but also for other purposes such as the features selection or features importance. By building models with different features, the accuracy/misclassification error and variable importance as well as covariance changes with them.

When selecting features, one can build SVMs, RF and other models with different features and compare the accuracy, which will change depending on the importance of each variable used to build the model. (Rakotomamonjy, 2003)

1. Initialization: $Ranked = []$; $Var = [1, \ldots, N]$
2. **repeat**
   (a) Train a SVM classifier with all the training data and the variables Var
   (b) **for** all variables in Var, **do** evaluate the ranking criterion $R_c(i)$ of variable i **endfor**
   (c) $best = \arg\min_i R_c$
   (d) rank the variable that minimizes R_c: $Ranked = [best\ Ranked]$;
   (e) remove the variable that minimizes R_c from the selected variables set: $Var = [1, \ldots, best-1, best+1, \ldots, N]$
3. **until** Var is not empty

Figure 1: Outline of the SVM-based feature selection algorithm.

Figure 8: example of feature selection using SVM, mentioned by Rakotomamonjy, 2003

(6) Feature selection

Before creating pretty algorithms that will make our life easier, there needs to be data. That data has to be cleaned and adapted before even considering any application. As the amounts of data increase, the computation becomes largely expensive, and discussions are held about the feature selection and feature importance.

According to Isabelle Guyon and André Elisseeff, noise and herewith class separation can be obtained by adding features, that are presumably redundant. Common methods for obtaining the importance of the features include covariance or correlation. However, as Guyon and Elisseeff state in their article on variable and feature selection that perfectly correlated variables are truly redundant in the sense that no additional information is gained by adding them. The same way, high variable correlation does not mean absence of variable complementarity. A variable that is completely useless on its own can provide significant performance improvement when added to other features. (Guyon & Elisseeff, 2003) It is therefore very difficult to assess, which features add value to the dataset.

According to Eugene Tuv et al., there are three major categories of feature selection methods:

- Filter methods, which usually considers the conditional probability distribution of the response at a predictor value.
- Wrapper method, which departs from the fact that the prediction accuracy of a model directly measures the value of a feature set.
- Embedded methods, which use all the variables to generate a model and then analyse the model to infer the importance of the variables. Consequently they directly link variable importance to the learner used to model the relationship (exp.: RF). (Tuv, Borisov, Runger, & Torkkola, 2009)

(a) Advantages of RF

The main advantage of using Random Forest for variable importance/feature selection is that it measures the importance of each predictor individually as well as in multivariate interactions with other variables. Its weakness is however the features importance measure, which is bias towards correlated variables. (Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008) According to Toloşi et al., *"weights of the features belonging to groups of correlated features decrease as the sizes of the groups increase, which leads to incorrect model interpretation and misleading feature ranking."* (Toloşi & Lengauer, 2011)

All of the above mentioned methods have problems as none of them can really tell which features are really important. There are too many factors defining the importance to be able to calculate the importance, which is the main issue.

ii) Machine Learning in the legal sector

We have seen it in the banking industry as well as in the media industry, in the medicine, there is hardly one sector that has not seen the power of Machine Learning. The legal sector, on the other side, has taken some time to accept it due to its traditionally conservative nature.

First steps came with Harold J. Spaeth, who provided significant insights into the case and political factors that helped drive the Court's decision making. After a long period of silence, the 2002 forecasting project by Andrew Martin, Kevin Quinn, Theodore Ruger and Pauline Kim, showed that machine outperformed human beings in the prediction of the outcomes of the US Supreme Court with a tree-based prediction model.

The progress continued with LexisNexis and Westlaw, which created a database containing huge amounts of case details, with the possibility to search through them. It was the basis for everything that followed, because once the data is digitalised, it opens up the doors for any analytics.

Today, the market is still slow but growing steadily. Machine Learning is used to make discovery easier, to draft or analyse contracts for clauses, for billing, compliance or for the prediction of the Court's outcomes.

Digital aggregation has been increasingly testing the boundaries of copyright law for the past decade, challenging the set decision by the US Supreme Court Feist Publications, Inc. v. Rural Telephone Service Co., Inc, which states that any spark of creativity *"no matter how crude, humble or obvious it might be"*, qualified an original work for copyright protection (Chiou & Tucker, 2016). This set in motion a need for machine learning solutions in the Intellectual Property (IP) law, which counts currently to the litigation, where case outcome prediction is developed the best. (Katz, 2016) One example of this machine learning application is Lex Machina, which *"mapped every electronically available patent litigation event and outcome to bring openness and transparency to IP law"*.

An article by McShane et al. reports about a fully predictive model of securities fraud class action law suits, which predicts the likelihood of settlement and the expected settlement amount. (McShane, Watson, Baker, & Griffith, 2012)

Another area, with continuous investment flow is the area of e-discovery. (Gabe, 2016) According to Katz, an average of 45-50% of civil litigation costs are attributed to discovery. (Katz, 2016)

Manzama.com contributes the area of discovery by sharing insights from news, saving time spent on searching for relevant information. Quislex.com then specialises in document review, legal spend and compliance management.

A blooming part of the advanced analytics market in the legal sector is analytics surrounding contracts. According to Harry Surden,

> *"[Contracts] aggregated across multiple parties, agreements and transactions, the net costs of conforming to a network of contractual obligations can be significant."* (Surden, 2016)

Kirasystems.com provides a software, for uncovering relevant information from contracts and relevant documents, and spotting hidden risks, while eBrevia.com provides a similar service as kirasystems.com, analysing 50+ documents for important information contained in contracts in less than a minute, saving hours of work and assuring at least 10% more accurate outcomes. Legalrobot.com goes even further by comparing the contract against industry standards and across jurisdictions, saving lots of valuable time.

Recommind.com then uses different machine learning techniques to visualise the terms of a contract and make it more understandable, giving the complex legal language may be difficult to understand for a non-lawyer.

Ethereum.org is a little bit more futuristic, using blockchain to run smart contracts that do not require any third party interference, avoid fraud, downtime as well as censorship.

Ravel Law on the other hand specialised in the judge's analytics, understanding the judges and their background, helping understand the risks coming with each judge based on his or her previous decisions.

A lot of work has been done in the area of procedural law (Engle, 2004), which is clear and determined, which makes it easy to use deductive reasoning. The same kind of reasoning used by judges. An example of that would be the syllogism:

All men are mortal, (major premise)

Socrates is a man, (minor premise)

Therefore, Socrates is mortal.

By giving the model all the rules (concepts), the model only has to apply the rules upon which it can make decisions without even being submitted to a judge as no further analysis is needed.

These few examples are a clear sign of Machine Learning disrupting the legal industry the way no one ever thought would be possible. Replacing lawyers with a machine learning algorithm the same way drivers are being replaced as we speak and doctors receive recommendations on the best medication for each patient, based on collected data.

Legal sector is one of the last sectors to face the consequences of disruption, surprisingly, although it is one of the last ones and also the slowest one to adopt changes to the law, the legal profession is accepting and even embracing machine intelligence as an input.

The American Bar Association has already categorised machine intelligence as *"either an inanimate tool, much like a computer, or as a non-lawyer assistant"* and

> *"American lawyers are not only allowed to use machine intelligence, they are required [to maintain the requisite knowledge and skill necessary to provide competent representation]"* (American Bar Association, 2016)

Furthermore,

> *"[A]lawyer should keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology."* (American Bar Association, 2016)

Although there are objections from lawyers that lawyers lack the requisite knowledge to supervise such vendor of machine learning solutions. Responsibilities regarding non lawyer assistance (in this case machine intelligence) acknowledge that lawyers may seek non-lawyer help outside the law firm when providing legal service to the client.

There has been progress but that does not exclude existing barriers. Legal work is not an assignable data object, and even though it is atomised, it represents unauthorised practice laws, which prohibit non-lawyers from providing personalised legal assistance, which excludes anyone from the legal sector, who does not hold a law degree.

This barrier is probably the main factor, reducing the power as well as competition of Machine Learning in the legal sector. Conforming to an article by John O. McGinnis and Russell G. Pearce, in practice however, the market for these services has become de facto deregulated. (McGinnis & Pearce, 2014)

A market, that is slowly but surely starting is in the prediction of the court's outcomes. A few prediction models were created in the United States, one of them figuring in the following section, but the prediction models are not only a phenomenon from the United States. There has also been some development in this area in the EU, with a model, predicting the outcomes of the European Court for Human Rights with a different method than the one created in the US, using Natural Language Processing. (Aletras, Tsarapatsanic, Preoţiuc-Pietro, & Lampo, 2016) This market is therefore definitely about to start growing as soon as the models get better prediction accuracy and the business world becomes aware of such tools.

(1) Prediction model by Daniel Martin Katz, Michael J Bommarito II, and Josh Blackman

The prediction model by Daniel Martin Katz, Michael J Bommarito II, and Josh Blackman is a fully predictive model.

> *"Using only data available prior to the date of decision, our model correctly identifies 69.7% of the Court's overall affirm / reverse decisions and correctly forecasts 70.9% of the votes of individual justices across 7,700 cases and more than 68,000 justice votes.(…) it is the first robust, generalized, and fully predictive model of Supreme Court voting behavior offered to date."* (Katz, Bommarito II, & Blackman, Predicting the Behavior of the Supreme Court of the United States: A General Approach, 2014)

The code for the model is publicly available and allows for closer analysis. It is written in Python/iPython, used for interactive programming in various programming languages, enabling the user to combine different functions and getting the best of all languages supported by iPython. It is to be run with Anaconda, which removes difficulties connected with numerous dependencies that are to be installed before running the code.

The model is a tree-based ensemble model, namely Extremely Randomized Trees, which was first introduced by Geurts et al. in 2006.

In order to predict the dependent variable, wide number of features was leveraged, including the features from the Supreme Court Database (Washington University Law, 2016), which mostly include the case facts, Segal Cover Scores (Segal, 2016), which measure the '*perceived qualifications and ideology'* and other features, developed with the help of feature engineering.

There are over 90 features (Appendix 2) analysed by the model, which finds the most important ones and uses them to build extremely randomised trees.

As mentioned in the Machine Learning section, there are three main methods for feature selection. This model benefits from the embedded method, which uses the learning part to select the most significant features. No filter is thus used for the determination of the most significant features. Embedded methods have usually higher capacity than filter methods and are consequently susceptible to overfit, it is therefore important to have a large training set to limit overfitting, as in the case of this model. The method consists of learning a model and evaluating the feature importance at the same time, depending on its contribution to a model. The higher the number of the weight, the higher the importance of the feature.

It is nevertheless important to note that many of the features are highly correlated, which makes clear interpretation quite a complex task because features belonging to groups of correlated features decrease as the sizes of the groups increase. Some important features risk therefore to be assigned a lower weight due to their correlation, which is very misleading.

This tree-based ensemble method does not rely on the bagging procedure as it is the case in RF. Instead, it uses the very same input set for training of all trees. Additionally, this method "*selects the split by indexing randomly across both, the variable index and variable splitting value."* Whereas RF select the best/optimal splitting condition. The idea behind extra trees is to reduce variance and the choice to use full original learning sample to minimize bias. (Katz, Bommarito II, & Blackman, Predicting the Behavior of the Supreme Court of the United States: A General Approach, 2014)

In order for the model to remain robust to temporal changes or changes in composition of the Court, extra trees are grown to reduce the variance and bias of the model.

For validation of the model, is used stratified k-fold cross-validation with 10 folds per training to make sure that there is no lucky draw and the validity of the model remains unaffected.

The prediction model has two parts (1) justice level vote prediction model and (2) prediction of the overall case outcome, which give more information about the decision outcome.

The whole model relies heavily on pre-processing and features engineering, see below:

```
# Calculate values
previous_direction = raw_data.loc[previous_index, "direction"].mean()
cumulative_direction = raw_data.loc[cumulative_index, "direction"].mean()
previous_action = raw_data.loc[previous_index, "justice_outcome_disposition"].mean()
cumulative_action = raw_data.loc[cumulative_index, "justice_outcome_disposition"].mean()
previous_agreement = (raw_data.loc[previous_index, "justice_outcome_disposition"] \
    == raw_data.loc[previous_index, "case_outcome_disposition"]).mean()
cumulative_agreement = (raw_data.loc[cumulative_index, "justice_outcome_disposition"] \
    == raw_data.loc[cumulative_index, "case_outcome_disposition"]).mean()

# Lookup or calculate values
if term in previous_court_direction_cache:
    previous_court_direction = previous_court_direction_cache[term]
    cumulative_court_direction = cumulative_court_direction_cache[term]
    previous_court_action = previous_court_action_cache[term]
    cumulative_court_action = cumulative_court_action_cache[term]
    previous_court_agreement = previous_court_agreement_cache[term]
    cumulative_court_agreement = cumulative_court_agreement_cache[term]
else:
    # Get the court-term masks
    previous_court_index = (raw_data.loc[:, "term"] == (term-1)) \
                           & (raw_data.loc[:, "justice_outcome_disposition"] >= 0)
    cumulative_court_index = (raw_data.loc[:, "term"] < term) \
                           & (raw_data.loc[:, "justice_outcome_disposition"] >= 0)

    # Calculate court direction
    previous_court_direction = previous_court_direction_cache[term] \
        = raw_data.loc[previous_court_index, "direction"].mean()
    cumulative_court_direction = cumulative_court_direction_cache[term] \
        = raw_data.loc[cumulative_court_index, "direction"].mean()
    # Calculate court action
    previous_court_action = previous_court_action_cache[term] \
        = raw_data.loc[previous_court_index, "justice_outcome_disposition"].mean()
    cumulative_court_action = cumulative_court_action_cache[term] \
        = raw_data.loc[cumulative_court_index, "justice_outcome_disposition"].mean()
```

Figure 9: Screenshot: feature engineering, code by Katz et al.

Before building the model, the dataset had to be cleaned and background knowledge was used to create new features, which capture the importance of the variable.

The model only relies on data available prior to the decision, however even that data has to be cleaned to provide valuable information. It herewith perfectly mimics the methodology used by the Supreme Court experts and probably even to the justices, which use the same logic, but have more information, than the model. Still, the model achieves better accuracy.

What the experts and justices do not do is to evaluate the importance of each variable and the prediction of each justice in percentage, calculating the prediction accuracy of each justice per term, see heat map in the Appendix 4. Human beings however do not assign lower weight to correlated variables, which presents in this case clear bias.

The model is perfectly designed to predict the outcomes of the US Supreme Court, ensuring generality as well as preventing against possible bias. It has however problems assigning correct weights to each predictor variable, which reduces its interpretability as well as its validity.

4) What can a machine learning algorithm tell us about the decision making of the US Supreme Court?

Without doubting the expertise of the US Supreme Court, it is clear that even the Supreme Court has certain limitations, which have to be addressed. By analysing the decision making, limitations of human decision making and the advantages of data driven decision making, it is clear that data driven decision making should be applied not even in the area of business, but also in the area of law, which would benefit from the interpretability as well as accuracy. The Court may be a little bit behind with the technology, as current Chief Justice Roberts says:

> *"Like other centuries-old institutions, courts may have practices that seem archaic and inefficient—and some are. But others rest on traditions that embody intangible wisdom.*
>
> *Judges and court executives are understandably circumspect in introducing change to a court system that works well until they are satisfied that they are introducing change for the good."* (Chief Justice Roberts, 2014)

The Court may move slowly, but as one could see, there has already been some progress on the business side as well as the legal part – the American Bar Association, and there is no doubt that more will follow. The first step could be the model presented in the case study – the prediction model by Katz et al., which already includes basics. It has difficulties assigning correct weight to correlated features, but as this part is only an additional product to forecasting the outcomes, which, it does very well, the assignment of weight is just a little issue. Nevertheless is the Chief Justice Roberts right to wait until the model is free of bias.

Although the validity of these findings is reduced because the importance of correlated features has a lower weight, the model shows that the justice's background is not as important as people believe it to be. According to the model, the importance of gender of the justices is only 0.00205, party of the President who appointed them only 0.00604 and the year of birth 0.00793, with the weight of justice's background in average at 0.04403.

According to the model, what actually matters are the case facts with an average weight of 0.22814. The variable with the highest rate being the month of the argument (0.02014), cert reason (0.01408), month of decision (0.01349), Lower Court Disposition (0.01125) and Lower Court Disagreement (0.00706).

These results seem perfectly reasonable given that the most controversial cases are scheduled for the end of the term. Cert reason – reason for granting petition for certiorari, which is often not given by the justices, but when they do give a reason, it has an impact on the decision making. Month of decision correlates closely with the month for argument, lower court disposition is in theory one of the most important reasons for review of the lower court decisions as it specifies the treatment whose decision the Supreme Court reviewed accorded the decision of the court it reviewed, shortly, whether the lower court's decision was affirm, reverse or remand. Finally, the Lower Court Disagreement, which informs about disparities in the interpretation of the law by telling the model, whether there were any dissents in the lower court's decision.

Although these findings need to be inspected with an improved model, which would not bias towards correlated variables, they are similar to the results from the Natural Language Processing method, used in the analysis of the European Court of Human Rights and deserve therefore more credit. Not only does the algorithm by Katz et al. predict the Court's outcomes, it also contributes to better understanding of the decision making of the Court. (Aletras, Tsarapatsanic, Preoţiuc-Pietro, & Lampo, 2016)

5) Why do we need Machine Learning on the Court?

The US Supreme Court as a common Law jurisdiction with application of stare decisis gets more power than any other institution would dream of given its authority of constitutional interpretation. Its Supremacy allows it to take action in the most controversial matters, where the Constitution does not provide any answers. Instead, the nine justices rely on their personal experience and personal beliefs, to which they are entitled considering they are completely independent in their judgment.

The plural membership should protect misuse of such independence and enhance diversity on the Court. The justices are however appointed by the President, who only appoints elites as one can see at the current Court and who makes his homework by looking into their background to ensure alignment of values. The political character of the Supreme Court then becomes very obvious to say the least.

The justices at the bench are appointed to a life service to achieve impartiality. They are however appointed by the President, to whom they often remain loyal, which is not the definition of impartial. This trend is difficult to miss when looking at the current Court (Appendix 1), which consists in majority of conservative and pro-business justices.

Another problem is the process of reviewing a case, and deciding about granting a petition for certiorari. Although the criteria for granting a petition is clear, only 1% of all filed cases are granted review. Such decision should be subject to analysis as it is a major part of the decision making of the US Supreme Court. Moreover, a major part of the selection process is done by law clerks, who do not even get appointed and might pursue their personal interests.

There are numerous factors playing a role in the decision making of the US Supreme Court, such as constitutional interpretation, personal beliefs, and the background of the justices. These parts of the decision making process can be partially influenced by the executive. According to Peter Drucker this stage is the most important decision, as this decision decides about the performance of the future group.

There are however other factors that are not predictable at the appointment stage, such as amicus briefs, filed in each case. The number of amicus briefs filed each time increases (out of the 275 business related cases filed in the Roberts Court, 261 cases included at least one amicus brief) while the number of amicus briefs read is possibly decreasing as some law clerks are instructed to filter amicus briefs, which were filed by unexperienced attorneys. Experienced attorneys are therefore in clear advantage.

A party with sufficient experience is the Government – the Solicitor General files amicus briefs very frequently; and not just that. He has the power to file amicus brief on any occasion and is often asked by the justices for opinion, suggesting his opinion matters. This argument is supported by the strong professional background of the sitting justices.

According to Appendix 5, out of 275 business cases, the Solicitor General filed an amicus brief on 150 occasions. The Solicitor General supported businesses in 72 cases, which is about 26% of all business cases and 48% of cases in which it filed amicus brief. The Solicitor General filed 115 amicus briefs in business related cases – the exact same amount as the Chamber of Commerce – showing they might have similar interests.

a) Increasing amounts of data improve Decision-making

Increasing the amounts of information reduces uncertainty and improves the decision making – in theory.

All information connected to a case is relevant, including all previous cases as stare decisis applies. This can become a huge task, only to review all the relevant cases not to mention other cases subject to discovery. Because this task is already impossible to manage, law firms keep investing into digital search engines for case information and e-discovery, which uses Machine Learning to analyse all relevant data to a case.

For human beings, it would take ages to find and analyse all the relevant information, especially due to the huge amount of information and the cognitive limitations. Humans are more interested in the story than in the numbers, which are more important in the decision making, making humans redundant and perfectly replaceable by a machine (Katz, 2016) with the benefit of better precision and accuracy of results.

As Kahneman explained, the best solution is then to include a machine into the decision-making. This method has been used across all different industries, because not only does it analyse all data in a short time span, it also improves performance, accuracy as well as understanding of decision-making. (Reeves & Ueda, 2016)

Law provides a lot of information which is already digitalised, and only requires a business application. Of course, legal industry is very traditional, however experience shows, that only in about half a dozen studies the clinical decision making achieves better results than a mechanical one. It further shows that the more information is available, the better results are. The opposite applies for human beings.

One cannot manage what cannot be measured and until now, the Court has been a black box. This is about to change with Machine Learning which can measure the decision making of the US Supreme Court. Following the trend from businesses applying data driven approach in their decision making, this could have a positive impact on its performance.

This could only be good, not only for the sake of the Court, but for the correct management of businesses. Litigation impacts the business environment and in order to move in the right direction, the understanding of its functioning needs to be understood. The Supreme Court's decisions are like a domino, what happens at the beginning has a ripple effect on all the other lower courts, even the smallest ones.

The question is: how can data points make me understand. By finding patterns and outliers in vast amounts of data, it finds information that would be overlooked or would never be found due to the amount of data. Some algorithms are so called black boxes, which only need to provide an accurate answer, which they do better than any other model. Other algorithms provide great interpretation and show exactly the process which led to a decision, such as decision trees.

The validity of the model can be measured as it is built using a training sample and the model is then tested on an unseen data and the amount of correct and wrong predictions is measured. It explains what variables were used and what variables were redundant, all automatically and without human bias.

As one can see in the example used, the prediction model by Katz et al. managed to predict the decisions of the Supreme Court to a **69.7%** accuracy, based on about 90 variables, moreover, it managed to assess the importance of each variable. The model only had about 90 variables, which is without any doubt less information than the legal experts, but the model as opposed to legal experts did not have any trouble forecasting the outcomes of the Court and attaching more or less correct weight to each feature in the model, see Appendix 2. It showed the most important factors in the decision making, which by themselves already improve the understanding of the decision making.

By adding more information to the model and removing correlation bias, the accuracy should improve, moreover, including all available information, the model should achieve nearly perfect prediction and we might end up discovering unknown patterns and factors, which are not even being considered by the legal experts as it was the case in pathology.

6) How can the understanding of the US Supreme Court be further improved?

According to Alexander Wissner-Gross, perhaps the most important news of our day is that datasets – not algorithms – might be the key limiting factor to development of human level AI. (Wissner-Gross, 2016) This statement may be true, but an algorithm still provides an important part of the solution.

The model by Katz, Bommarito II, & Blackman suffers from correlation bias, which makes the determination of features importance difficult. A reduction of correlation bias is therefore needed to enable a better understanding of the Court's decision. A possible solution would then be for example hierarchical clustering of the features as suggested by Laura Toloşi and Thomas Lengauer. (Toloşi & Lengauer, 2011)

Further, decision making literature suggests, the decision making is improved through the reduction of uncertainty, which happens by adding more information to a dataset. Prioritising and cultivation of new datasets is thus essential for the improvement of the decision making.

Literature reviewed suggests that important features can improve the accuracy of a prediction model and together with that – the decision making. According to Lee Epstein and other legal scholars, amicus briefs are important in the decision making of the US Supreme Court, which is why this feature should be added to the dataset. Moreover, it suggests that not only does it matter whether an amicus brief was filed, it is important to know who filed the amicus brief. The justices have clear preferences as to the attorney filing the brief, which results in clear bias, making the identity of the filing entity an important feature. As mentioned previously the US Solicitor General and the Chamber of Commerce have the relevant experience and herewith the relevant authority at the Court.

Additionally, as the Solicitor General tends to support businesses and the justices tend to support Solicitor General's opinion, a case, where the Chamber of Commerce and Solicitor General filed an amicus brief in favour of business have a higher chance of being pro-business than not, which is why the following variables should be included in the dataset. They should be added because it is believed that their presence would add value to the dataset and herewith to the accuracy of the prediction model in which it would be used.

(1) Do the added features add value?

As mentioned in the Machine Learning section, additional features can add value, but they can also add noise. A variable can be completely useless on its own and completely indispensable in a dataset surrounded with other variables, which makes it very hard to measure such value. There are different ways to measure the value, such as filter methods, wrapper methods, and embedded methods

The wrapper and embedded methods were used to measure the value of the added features. The wrapper method was done using SVM and the embedded method was done using RF. All machine learning methods were done using r in rStudio. The target variable is binary, consequently SVM as well as RF were used as classification task.

(a) Description of the dataset and pre-processing

The original dataset was in a csv file, and included 8737 observations and 53 variables. As this research is mainly about the Roberts Court, the dataset was filtered and only Roberts Court cases were left. Due to the fact that in the Chief Justice variable, there was only one value left, the variable was no longer meaningful and was therefore deleted from the dataset (leaving 869 observations with 52 variables).

The cases were supposed to be business oriented, meaning that the left 869 cases were then skimmed through and only the relevant cases were left in the dataset (275 observations), leaving only about 3% of the cases for further analysis.

Following that, additional features were added- class- whether the case was affirmed or reversed, in order to be able to do supervised learning. The additional features were amicus briefs- whether an amicus brief was filed for the given case. Whether the amicus brief was from the United States Solicitor General or from the Chamber of Commerce and in case there was an amicus brief from the Solicitor General, whether it was pro-business.

Added feature	Definition
Class **'class'**	1 affirmed, 0 reversed
Amicus Briefs 'Amicus.Brief'	1 filed, 0 not filed
Amicus Brief for the Chamber of Commerce 'Amicus.Brief.from.the.Chamber.of.Commerce'	1 filed, 0 not filed
Amicus Brief from the United States 'Amicus**.Brief.from.the.United.States'**	1 filed, 0 not filed
United States supporting business 'US.supporting.Business'	1 yes supporting business, 0 not supporting business

Figure 10: Added features

Finally, some variables had to be removed as they do not add any information to the dataset because they serve solely as identification for each observation (caseId, docketId, caseIssuesId, voteId, lexisCite, docket, caseName, usCite, sctCite,). Some variables were removed as their value is constant and therefore do not add any value (splitVote, caseDispositionUnusual, voteUnclear, jurisdiction). The reason why their value is constant is because the information they contain is the same for each of the observations. If all observations were used, the variables would have added value because the value of the variable changed, however under the Chief Justice, whose cases were selected, the value in the variables splitVote, caseDispositionUnusual, and voteUnclear remained the same.

The splitVote variable indicated whether the vote variables pertain to the first or second issue. As such events are very rare, 99% of all observations have a ‘1’ as a value. caseDispositionUnusual signifies that the Court made an unusual disposition of the cited case, which already signifies the rarity of this event. voteUnclear means that the votes of one or more justices is unknown, which according to the code book of the dataset used happened mostly under Chief Justice Marshall. jurisdiction has value ‘1’ for all cases, because the cases jurisdiction was accepted based on writ certiorari, which is the most common value for business cases (Washington University Law, 2016)

Additionally, some variables (numberOfJustices, winningParty, dateArgument, petitionerState, respondentState, adminAction, adminActionState, caseOriginState, caseSourceState, authorityDecision2, lawMinor) were removed because they contained too many missing values, which would affect the modelling.

The dataset contained some dates, which were in characters, and therefore could not be used in many machine learning models, which mostly require numeric values. Therefore they were converted into date values. The two variables were the dateDecision and dateArgument. Due to the difficulty to incorporate the two variables into any model, the delay between these two variables was calculated and added as a new variable, while the former date variables were deleted. The dataset created can be found in Appendix 5. The full dataset is referred to as ‘*data*’, and the dataset without the added values is referred to as ‘*modified’*.

As one can see in the Appendix 6, none of the methods show any improvement in accuracy. The reason for that may be the same reason as in the case of the prediction model by Katz et al. – Correlation.

If numerical features correlate, they do not add any additional information to classical machine learning methods as the information is determined by the correlating features. According to Toloşi & Lengauer, 2011, *“correlated features are used interchangibly in the decision trees of random forests”*and receive smaller weights due to shared responsibility in the model. This is a problem, because the added features are all correlated.

RF is a great method as it calculates the importance of each predictor individually as well as in multivariate interactions with other variables. However when the subset of features includes correlated features, it is bias. The weight of correlated features decreases with the size of the group of correlated features. (Toloşi & Lengauer, 2011). Features, which are redundant to the dataset are then boosted and receive higher weight whereas important features, which are correlated receive lower weight due to correlation, making the interpretation much harder. If this group of correlated features is big enough, the correlated features then appear completely irrelevant, due to reduced weight. (Toloşi & Lengauer, 2011)

The results from the RF are not completely useless though. They still show the importance of each of the features, including the added features, where Amicus brief from the United States receive a weight of 0.5138, Amicus brief from Chamber of Commerce 0.3422, Amicus brief supporting businesses 0.2376 and Amicus brief 0.0419, which in comparison to the most important variables such as caseDisposition or partyWinning appear to be completely insignificant. Nevertheless it is important to remember that their weight is significantly lower because of their correlation.

SVM and RF were then used to build models with one, two, three and four of the added variables to see whether the different features have different impact on the accuracy of the model.

In SVM, no change was observed. The results from RF however show that each variable has a different impact on the accuracy of the model, moreover even the combination of features has an impact on the accuracy and out-of-bag error.

The table below shows the out-of-bag errors for each of the variables present in the dataset.

Amicus Brief from the Chamber of Commerce Amicus Brief from the Solicitor General United States Supporting business 1.64	Amicus Brief from the Solicitor General United States Supporting business 1.23	United states Supporting business 1.64
Amicus Brief from the Solicitor General United States Supporting Business Amicus Brief 1.64	United States Supporting Business Amicus Brief 1.23	Amicus Brief 0.82
United States Supporting business Amicus Briefs Amicus Brief from the Chamber of Commerce 1.23	Amicus Briefs Amicus Brief from the Chamber of Commerce 1.64	Amicus Brief from the Chamber of Commerce 1.23
Amicus Brief Amicus Brief from the Chamber of Commerce Amicus Brief from the Solicitor General 0.82	Amicus Brief from the Chamber of Commerce Amicus Brief from the Solicitor General 1.23	Amicus Brief from the Solicitor General 0.82

7) What does Machine Learning mean for the Court and for Businesses?

The court is increasingly becoming overwhelmed by the amount of information, which cannot be processed by human beings. The amount of information is just increasing exponentially and the only way to take all the information available into account is to use Machine Learning to process it and use its outcomes to guide us in our decision making by giving us objective insights, relating to the case.

The existing applications already help us with e-discovery and information retrieval, which saves hours and hours of work and avoids loosing crucial information in the huge amounts of information. Additionally it adds efficiency. This enables currently attorneys to make informed decisions, about future disputes coming from lack of information.

The case example then enables us to better understand what is behind decision making by the US Supreme Court. Not only does it provide a way to predict the outcomes of the US Supreme Court it also includes the embedded feature selection method, which tells us what are the most important features in the model and herewith in the decision making of the US Supreme Court. Although this feature selection method needs to be improved to reduce bias towards correlated variables, the fixation of this problem is just a question of time. There are numerous papers discussing possible ways to reduce this bias, by for example combining supervised learning with unsupervised learning (Park, Hastie, & Tibshirani, 2006) or different computation of the variable importance measure. (Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008)

Once this modification is in place the model will probably achieve better results with more relevant data, however until then, it is hard to measure such improvement because of the standing bias towards correlated features. As Alexander Wissner-Gross says, datasets are probably more important in the development of AI, however first, the correct algorithms have to be in place because without an algorithm, the datasets will not be used.

For that, legal experts are needed to assist analysts and share their knowledge, so that a dataset with all relevant information can be created and used to provide useful insights and knowledge.

a) Review of the US Supreme Court

The structure of the judicial branch did not take into account that a review of the Supreme Court will be needed. However with the increasing amount of controversial issues ending at the Court, where the Constitution does not provide any answer, the review becomes very important. Not only to make people understand the decision making, but also as a means to manage the judicial power.

In order to manage something, it has to be measurable, which is why everything becomes digitalised. It is not to be seen as a control measure but as a performance measure. All industries are heading this way, why should legal sector stay behind.

By the adaptation of data driven decision making, the performance improves and so does the interpretability of decision making, which is currently lacking. Moreover, three problems will be solved, resulting in increased confidence, performance and objectivity.

b) Advantages for Businesses

The advantages of prediction models of the US Supreme Court decisions are far-reaching. Apart from interpretability of each variable contributing to the decision making and possibility of missing legal review, it could improve decision making of the businesses.

> *"[l]aw affects the market and market players, but market players also affect the law and the way it is interpreted, applied, and changed over time. Thus, law is not just a static external force acting upon managers and their firms. Instead, law and organizations are "endogenously coevolutionary." By lobbying legislators and members of the executive branch, forming coalitions, and working directly with regulatory bodies, managers can help shape the environment in which they do business."* (Bagley, 2012)

The Supreme Court decisions are to be followed by the rest of the courts in the country, once the decision is delivered, all pending lawsuits involving the same legal provision are to apply the same results, making business cases a relevant factor in business decision making.

Knowing the outcome of the decision before the decision itself can save lots of time, money as well as reputation. Anticipation of the predicted negative decision of the US Supreme Court then means, the company will be able to settle the case before the negative outcome is announced.

As a result, the company saves a lot in attorney fees, Court fees, and time and does not loose clients because of a negative outcome. Moreover, by getting this information before the official announcement, it can invest money in a strategy that would remove their liability in face of another lawsuit regarding the same legal issue.

Looking into the most important factor, one could go even further, and invest in the factor that matters the most and changing the scenario to a positive one, where the company wins.

Another scenario would be a company that is planning on requesting a review from the US Supreme Court. Knowing, what are the most important factors in the decision making, the company can include these factors into the litigation strategy, improving their chance of a positive outcome.

Law firms would also benefit from such model, because they often receive a percentage of the outcome. By knowing up front, which cases are to be successful, they can save a lot of time, looking into the background of the case, by applying a simple machine learning model.

They could also use it for hiring purposes. Adding other variables, such as the name of attorney into the model, one could find out, which attorneys have a higher chance of winning or getting a case granted petition for certiorari.

Moreover, by knowing the value of the attorney, they can incorporate this into their pricing model. This then in return adds more transparency into the legal fees.

Finally, the legal profession does not currently allow for review of law firms, however, by incorporating them into the factors in a decision making, would allow for a measurement of their expertise, allowing for increased transparency in the spending and their charges as well as boost in competition.

8) Concluding remarks

In using Machine Learning for better understanding, one not only gets a better understanding of the Court whose processes remained secret for so long. One gets the missing review of the judicial branch, which is missing at this stage.

By creating a database with all the information about previously decided cases as well as the information about each of the justices, new features can be created based on the context. This modified database can then be used in a prediction model like the one created by Katz, Bommarito II, & Blackman, which not only makes conclusions on the probability of a positive or negative outcome, one gets even the variables, which are the most important in such decision. In this direction however, further research is needed to reduce bias against correlated features, which is present in the prediction model by Katz, Bommarito II, & Blackman.

If more relevant information is added, not only will it improve the accuracy of the model, it will also enable better understanding of the decision making.

For that it is however crucial that the algorithm does not discriminate against correlated variables as at the Supreme Court, a significant percentage is correlated. This requirement is not negotiable in this context as the features importance is in the heart of this problem. Moreover, before adding different variables to the dataset, their importance has to be measured to ensure an improvement and avoid noise. In this matter, legal expert will be crucial, as context is the key in any area of Machine Learning.

As mentioned in the previous chapters, data driven analysis of the decision making of the US Supreme Court would not be just good for protection against potential misuse, it can also be used as an element of decision making in the business environment, which is moved by the Supreme Court decisions, and changes to the business strategy and decisions always need to be made in light of its decisions.

Finally, as with other applications of data driven decision making, although there is so far no evidence suggesting this trend, using a predictive model to anticipate the decisions of the United States Supreme Court, businesses will surely become more profitable and achieve better performance through risk reduction and damage anticipation.

9) Works cited

Aletras, N., Tsarapatsanic, D., Preoţiuc-Pietro, D., & Lampo, V. (2016). *Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective.*

American Bar Association. (1. October 2016). *American Bar Association*. http://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_1_1_competence/comment_on_rule_1_1.html

Bachmann, K. H. (2011). *Behavioural Finance for Private Banking.* Wiley.

Bagley, C. E. (2012). *Managers and the Legal Environment: Strategies for the 21st Century.* Cengage Learning.

Barrett, P. (2015). Supreme Court Sides With Big Business (Again). *Bloomberg*.

Beck, A. H., Sangoi, A. R., Leung, S., Marinelli, R. J., & Nielsen, T. O. (18. November 2016). *Stanford Medicine*. http://med.stanford.edu/labs/vanderijn-west/documents/108ra113.full.pdf

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (27. July 1992). A Training Algorithm for Optimal Margin Classifiers. *Proceeding COLT '92 Proceedings of the fifth annual workshop on Computational learning theory*, S. 144-152.

Breiman, L. (2001). Random Forests. *Machine Learning* , 5-32.

Bryan, A. C. (1. October 2016). *University of Minnesota*. http://www.tc.umn.edu/~bryan202/AmandaBryanWebsite/Research_files/ClerksPaper.pdf

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (1. November 2016). *ssrn.com*. ssrn.com: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1819486

Buchanan, L., & O'Connell, A. (2006). A brief history of decision making. *Harvard Business Review*.

Chi, C.-L. (2009). *Iowa Research Online.* http://ir.uiowa.edu/cgi/viewcontent.cgi?article=1469&context=etd

Chief Justice Roberts. (2014). *Supreme Court*. https://www.supremecourt.gov/publicinfo/year-end/2014year-endreport.pdf

Chiou, L., & Tucker, C. (19. November 2016). *semanticscholar.org*. https://pdfs.semanticscholar.org/aad5/93f0bda7f3d990f9fedff7cc136cdf98b929.pdf

Cohen, B. P. (1991). *Developing sociological knowledge: Theory and method (2nd edition).* Chicago: Nelson-Hall.

Dahl, R. A. (1. October 2016). *heinonline*. http://heinonline.org/HOL/LandingPage?handle=hein.journals/emlj6&div=20&id=&page=

Engle, E. A. (2004). An Introduction to Artificial Intelligence and Legal Reasoning. *Richmond Journal of Law & Technology* .

Epstein, L., & Posner, E. A. (1. November 2016). *epstein.wustl.edu.* http://epstein.wustl.edu/research/JusticePresident.pdf

Epstein, L., Landes, W. M., & Richard, A. (1. October 2016). *minnesotalawreview.org*. http://www.minnesotalawreview.org/wp-content/uploads/2013/04/EpsteinLanderPosner_MLR.pdf

Epstein, L., Martin, A. D., Quinn, K. M., & Segal, J. A. (1. January 2007). Ideological Drift among Supreme Court Justices: Who, When, and How Important. *Northwestern University Law Review*.

Feldman, A. (1. October 2016). *Empirical Scotus*. https://empiricalscotus.com/2016/09/15/supreme-court-work-patterns/

Gabe, F. (2015). Survey: Some Lawyers View Data As Threat, Investment Growing. *Bloomberg-Big Law Business*.

Gabe, F. (2016). Andreessen Horowitz Invests $8.1M in E-Discovery and Other Tech News. *Bloomberg law-Big Law Business*.

Geurts, P., Ernst, D., & Wehenkel, L. (15. November 2016). Extremely randomized trees. *Machine Learning*, S. 3-42. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.65.7485&rep=rep1&type=pdf

Greenhouse, L. (2016). The Choice Confronting the Supreme Court's Chief Justice. *The New York Times*, SR5.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (19. Monat 2016). *The Affective Neuroscience Laboratory*. http://zaldlab.psy.vanderbilt.edu/resources/wmg00pa.pdf

Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research 3*, 1157-1182.

Harper, B. D. (1. October 2016). *The Penn Law Legal Scholarship Repository*. http://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1540&context=jcl

Hawley, J. D. (2012). The Most Dangerous Branch. *National Affairs*.

Hazard, G. C. (1978). Supreme Court as a Legislature. *Cornell Law Review*.

Ho, T. K. (14. August 1995). Random Decision Forests. *Proceeding ICDAR '95 Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1) - Volume 1*, S. 278.

Howe, A. (2015). Interpreting the Supreme Court: Finding Meaningin the Justices Personal Experiences. *Florida Law Review*.

Hughes, C. J. (21. November 2016). *Supreme Court*. https://www.supremecourt.gov/about/constitutional.aspx

Judex.com. (1. November 2016). *Judex.com*. http://www.judex.dk/medical-informatics/decision-support-systems.html

Kahneman, D., Rosenfield, A. M., Gaudhi, L., & Blaser, T. (2016). Noise: How To Overcome The High, Hidden Cost of Inconsistent Decision Making? *Harvard Business Review*. https://hbr.org/2016/10/noise

Kaplan, A. (1964). *The Conduct of inquiry: Methodology for behavioral Science.* San Francisco: Chandler.

Katz, D. M. (2016). Quantitative Legal Prediction—or—How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry. *Emory Law Journal*.

Katz, D. M., Bommarito II, M. J., & Blackman, J. (2014). *Predicting the Behavior of the Supreme Court of the United States: A General Approach.*

Langášek, T. (2011). *Ústavní soud Československé republiky a jeho osudy v letech 1920-1948.* Pilsen: Vydavatelství a nakladatelství Aleš Čeněk, s.r.o.

Law Cornell. (1. October 2016). https://www.law.cornell.edu/constitution/articleiii

Law, W. U. *The Genesis of the Database.*

Liptak, A. (2008). A Second Justice Opts out of a Longtime custom: The 'Cert.Pool'. *The New York Times*, A21.

Lynham, S. A. (19. November 2016). *SAGE Publications*. http://lepo.it.da.ut.ee/~cect/teoreetilised%20seminarid_2009%20sügis/2_seminar_KULTUURITEOORIA_10.11.2009/Method%20of%20theory%20building%20research%202002.pdf

Marr, B. (2016). How Big Data Is Disrupting Law Firms And The Legal Profession. *Forbes.com*.

Martin, A., Quinn, K., Ruger, T., & Kim, P. (2004). The Supreme Court Forecasting Project: Legal and Political Political Science Approaches to Predicting Supreme Court Decisionmaking. *Berkeley Law Scholarship Repository*.

McAfee, A. (2003). Big Data's Biggest Challenge? Convincing People NOT to Trust Their Judgment. *Harvard Business Review*.

McAfee, A. (2014). When Human Judgment Works Well, and When Iit Doesn't. *Harvard Business Review*.

McAfee, A. (18. November 2016). *Andrew McAfee*. http://andrewmcafee.org/2011/12/using-computers-to-cure-medicines-pathologies/

McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review*.

McCarthy, J. (2015). *Disapproval of Supreme Court Edges to New High.*

McGinnis, J. O., & Pearce, R. G. (2014). THE GREAT DISRUPTION: HOW MACHINE INTELLIGENCE WILL TRANSFORM THE ROLE OF LAWYERS IN THE DELIVERY OF LEGAL SERVICES. *Fordham Law Review*.

McShane, B. B., Watson, O. P., Baker, T., & Griffith, S. J. (2012). Predicting Securities Fraud Settlements and Amounts: A Hierarchical Bayesian Model of Federal Securities Class Action Lawsuits. *Journal of Empirical Legal Studies*.

Miller, T. W. (2014). *Modeling Techniques in Predictive Analytics with Python & R.* Pearson Education.

Nutt, P. C., & Wilson, D. C. (2010). *Handbook of Decision Making.* Wiley-Blackwell.

Park, M. Y., Hastie, T., & Tibshirani, R. (8. May 2006). Averaged gene expressions for regression. *Biostatistics*, S. 212-27.

Posner, E. (2015). The Supreme Court's Loss of Prestige. *Slate Magazine*.

Posner, R. A. (1986). Legal Formalism, Legal Realism, and the Interpretation of Statutes and the Constitution. *Case Western Reserve Law Review*.

Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking.* O'Reilly Media.

Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning*, S. 81-106. http://hunch.net/~coms-4771/quinlan.pdf

Rakotomamonjy, A. (2003). Variable Selection Using SVM-based Criteria. *Journal of Machine Learning Research 3*.

Reeves, M., & Ueda, D. (2016). Designing the Machines That Will Design Strategy. *Harvard Business Review*.

Salesforce. (1. October 2016). *Salesforce*. https://www.salesforce.com/form/pdf/ai-for-crm.jsp

Saunders, M. N., Thornhill, A., & Lewis, P. (2009). *Research Methods for Business Students (5th Edition).* Pearson Education Limited.

Scheiber, N. (2016). As Americans Take Up Populism, the Supreme Court Embraces Business. *The New York Times*, B1.

Segal, J. (1. November 2016). *Stony Brook University.* http://www.stonybrook.edu/polsci/jsegal/

Segal, J. A., & Cover, A. D. (1989). *Ideological Values and the Votes of U.S. Supreme Court Justices.*

Spriggs, J. F., & Wahlbeck, P. J. (1. October 2016). *Political Research Quarterly*. http://prq.sagepub.com/content/50/2/365.abstract

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (11. July 2008). Conditional variable importance for random forests. *BMC Bioinformatics*.

Surden, H. (1. November 2016). *Computable Contracts.* http://lawreview.law.ucdavis.edu/issues/46/2/articles/46-2_surden.pdf

The New York Times. (2012). Trust and the Supreme Court. *The New York Times*, A18.

The US Supreme Court. (1. October 2016). *Supreme Court of the Unites States*. https://www.supremecourt.gov/faq.aspx

Toloşi, L., & Lengauer, T. (10. May 2011). Classification with correlated features: unreliability of feature ranking and solutions. *Bioinformatics*.

Toobin, J. (2008). *The Nine, Inside The Secret World of the Supreme Court.* New York: Anchor Books.

Trochim, W. M. (1. October 2016). *socialresearchmethods*. http://www.socialresearchmethods.net/kb/

Tuv, E., Borisov, A., Runger, G., & Torkkola, K. (2009). Feature Selection with Ensembles, Artificial Variables, and. *Journal of Machine Learning Research*.

Washington University Law. (7. October 2016). *The Genesis of the Database*. http://scdb.wustl.edu/about.php

Wissner-Gross, A. (1. November 2016). *Edge.org*. https://www.edge.org/response-detail/26587

Witten, J. G., Hastie, D., & Tibshirani, R. (2013). *An Introduction to Statistical Learning.* New York: Springer.

Wladawsky-Berger, I. (19. November 2016). *MIT Initiative on the Digital Economy*. http://ide.mit.edu/news-events-media/blog/learning-apply-data-science-business-problems

10) Appendices

a) Appendix 1

<u>Justices currently sitting at the bench</u>

Chief Justice John Roberts (born January 27, 1955), appointed in 2005 to replace Chief Justice William Rehnquist. He was educated at the Harvard Law School, is known to have conservative judicial philosophy, belongs to Roman Catholic Church and is Republican, which made him a great nominee for the conservative President, George W. Bush. Prior to his appointment, he represented 19 States in the case United States v. Microsoft, he was also judge at the DC Circuit.

Anthony Kennedy (born July 23, 1936), appointed in 1988 to replace Lewis Powell. He was educated in Stanford and Harvard. Kennedy belongs to the Roman Catholic Church, is against gun control and for capital punishment, making him the perfect Republican. He was appointed by President Ronald Reagan. He is the most senior Associate Justice currently serving at the Court and is considered the swing vote on many of the tight decisions.

Clarence Thomas (born June 23, 1948), appointed in 1991 to replace Thurgood Marshall. He was educated in Holy Cross and Yale, he belongs to the Roman Catholic Church and is considered the most conservative Justice on the current Court. He favours originalism, is against gun control and against abortion, which makes him again, the perfect nominee for President George W. Bush.

Ruth Bader Ginsburg (born March 15, 1933), appointed in 1993 to replace Justice Byron White. She is the second female justice on the Court after Sandra Day O'Connor. She was educated at Cornell and Columbia. She is Jewish, the liberal wing of the Court. Before her appointment, she was an advocate of the advancement of women's rights as constitutional principle. She believes that the US law should be shaped by foreign laws and norms. She is a democrat and was appointed by the President Bill Clinton.

Stephen Breyer (born August 15, 1938), appointed in 1994 to replace Justice Harry Blackmun. Breyer was educated in Stanford and Harvard. He is Jewish like the Justice Ginsburg, appointed by Bill Clinton and therefore also Democrat. He is well known for his pragmatic approach to constitutional law, which appears to be more liberal. Breyer is specialised in administrative law and before joining the Court, he was special assistant to the US Assistant Attorney General for Antitrust.

Samuel Alito (born April 1, 1950), appointed in 2006 to replace Justice Sandra Day O'Connor. He was educated in Princeton and Yale, he belongs to the most conservative Justices on the Court, is the first Italian American on the Court, he belongs to the Roman Catholic Church and is Republican, as well as the other nominee for the conservative President, George W. Bush. Prior to his appointment, he prosecuted many cases that involved drug trafficking & organized crime.

Sonia Sotomayor (born June 25, 1954), appointed in 2009 to replace Justice David Souter. She was educated in Princeton and Yale. She belongs like the majority of this Court to the Roman Catholic Church. Prior to the Supreme Court, she was at the US District Court for the Southern District of New York, US Court of Appeals for the Second Circuit. She had been democrat until 2009, since then she is Independent. As the year of appointment shows, she was appointed by President Barack Obama.

Elena Kagan (born April 28, 1960), appointed in 2010 by President Barack Obama to replace Justice John Paul Stevens. She is the youngest Justice currently serving at the Court. She was educated at Princeton and Harvard, she is conservative Jewish and a Democrat. She is the only one on the Court, who is not married and has no children. Before joining the Court, she was the first Solicitor General of the United States, however she never argued a case.

b) Appendix 2

Justice and Court Background Information
Justice [S]
Justice Gender [FE]
Is Chief [FE]
Party President [FE]
Natural Court [S]
Segal Cover Score [SC]
Year of Birth [FE]

Case Information
Admin Action [S]
Case Origin [S]
Case Origin Circuit [S]
Case Source [S]
Case Source Circuit [S]
Law Type [S]
Lower Court Disposition Direction [S]
Lower Court Disposition [S]
Lower Court Disagreement [S]
Issue [S]
Issue Area [S]
Jurisdiction Manner [S]
Month Argument [FE]
Month Decision [FE]
Petitioner [S]
Petitioner Binned [FE]
Respondent [S]
Respondent Binned [FE]
Cert Reason [S]

Overall Historic Supreme Court Trends
Mean Court Direction [FE]
Mean Court Direction 10 [FE]
Mean Court Direction Issue [FE]
Mean Court Direction Issue 10 [FE]
Mean Court Direction Petitioner [FE]
Mean Court Direction Petitioner 10 [FE]
Mean Court Direction Respondent [FE]
Mean Court Direction Respondent 10 [FE]
Mean Court Direction Circuit Origin [FE]
Mean Court Direction Circuit Origin 10 [FE]
Mean Court Direction Circuit Source [FE]
Mean Court Direction Circuit Source 10 [FE]

Lower Court Trends
Mean Lower Court Direction Circuit Source [FE]
Mean Lower Court Direction Circuit Source 10 [FE]
Mean Lower Court Direction Issue [FE]
Mean Lower Court Direction Issue 10 [FE]
Mean Lower Court Direction Petitioner [FE]
Mean Lower Court Direction Petitioner 10 [FE]
Mean Lower Court Direction Respondent [FE]
Mean Lower Court Direction Respondent 10 [FE]

Current Supreme Court Trends
Mean Agreement Level of Current Court [FE]
Std. Dev. of Agreement Level of Current Court [FE]
Mean Current Court Direction Circuit Origin [FE]
Std. Dev. Current Court Direction Circuit Origin [FE]
Mean Current Court Direction Circuit Source [FE]
Std. Dev. Current Court Direction Circuit Source [FE]
Mean Current Court Direction Issue [FE]
Z-Score Current Court Direction Issue [FE]
Std. Dev. Current Court Direction Issue [FE]
Mean Current Court Direction [FE]
Std. Dev. Current Court Direction [FE]
Mean Current Court Direction Petitioner [FE]
Std. Dev. Current Court Direction Petitioner [FE]
Mean Current Court Direction Respondent [FE]
Std. Dev. Current Court Direction Respondent [FE]

Individual Supreme Court Justice Trends
Mean Justice Direction [FE]
Mean Justice Direction 10 [FE]
Mean Justice Direction Z Score [FE]
Mean Justice Direction Petitioner [FE]
Mean Justice Direction Petitioner 10 [FE]
Mean Justice Direction Respondent [FE]
Mean Justice Direction Respondent 10 [FE]
Mean Justice Direction for Circuit Origin [FE]
Mean Justice Direction for Circuit Origin 10 [FE]
Mean Justice Direction for Circuit Source [FE]
Mean Justice Direction for Circuit Source 10 [FE]
Mean Justice Direction by Issue [FE]
Mean Justice Direction by Issue 10 [FE]
Mean Justice Direction by Issue Z Score [FE]

Differences in Trends
Difference Justice Court Direction [FE]
Abs. Difference Justice Court Direction [FE]
Difference Justice Court Direction Issue [FE]
Abs. Difference Justice Court Direction Issue [FE]
Z Score Difference Justice Court Direction Issue [FE]
Difference Justice Court Direction Petitioner [FE]
Abs. Difference Justice Court Direction Petitioner [FE]
Difference Justice Court Direction Respondent [FE]
Abs. Difference Justice Court Direction Respondent [FE]
Z Score Justice Court Direction Difference [FE]
Justice Lower Court Direction Difference [FE]
Justice Lower Court Direction Abs. Difference [FE]
Justice Lower Court Direction Z Score [FE]
Z Score Justice Lower Court Direction Difference [FE]
Agreement of Justice with Majority [FE]
Agreement of Justice with Majority 10 [FE]
Difference Court and Lower Ct Direction [FE]
Abs. Difference Court and Lower Ct Direction [FE]
Z-Score Difference Court and Lower Ct Direction [FE]
Z-Score Abs. Difference Court and Lower Ct Direction [FE]

Figure 1. Variables Employed by the Model

Figure 11: Predicting the Behavior of the Supreme Court of the United States: A General Approach

c) Appendix 3

Justice and Court Background Information	
Justice [S]	0.00781
Justice Gender [FE]	0.00205
Is Chief [FE]	0.00283
Party President [FE]	0.00604
Natural Court [S]	0.00764
Segal Cover Score [SC]	0.00971
Year of Birth [FE]	0.00793
TOTAL	**0.04403**

Case Information	
Admin Action [S]	0.00978
Case Origin [S]	0.00971
Case Origin Circuit [S]	0.00845
Case Source [S]	0.00953
Case Source Circuit [S]	0.01015
Law Type [S]	0.01370
Lower Court Disposition Direction [S]	0.01190
Lower Court Disposition [S]	0.01125
Lower Court Disagreement [S]	0.00706
Issue [S]	0.01541
Issue Area [S]	0.01469
Jurisdiction Manner [S]	0.00595
Month Argument [FE]	0.02014
Month Decision [FE]	0.01349
Petitioner [S]	0.01406
Petitioner Binned [FE]	0.01199
Respondent [S]	0.01490
Respondent Binned [FE]	0.01179
Cert Reason [S]	0.01408
TOTAL	**0.22814**

Overall Historic Supreme Court Trends	
Mean Court Direction [FE]	0.00988
Mean Court Direction 10 [FE]	0.01997
Mean Court Direction Issue [FE]	0.01546
Mean Court Direction Issue 10 [FE]	0.00938
Mean Court Direction Petitioner [FE]	0.00863
Mean Court Direction Petitioner 10 [FE]	0.00904
Mean Court Direction Respondent [FE]	0.00875
Mean Court Direction Respondent 10 [FE]	0.00925
Mean Court Direction Circuit Origin [FE]	0.00791
Mean Court Direction Circuit Origin 10 [FE]	0.00864
Mean Court Direction Circuit Source [FE]	0.00951
Mean Court Direction Circuit Source 10 [FE]	0.01017
TOTAL	**0.12663**

Lower Court Trends	
Mean Lower Court Direction Circuit Source [FE]	0.00962
Mean Lower Court Direction Circuit Source 10 [FE]	0.01017
Mean Lower Court Direction Issue [FE]	0.01334
Mean Lower Court Direction Issue 10 [FE]	0.00933
Mean Lower Court Direction Petitioner [FE]	0.00949
Mean Lower Court Direction Petitioner 10 [FE]	0.00874
Mean Lower Court Direction Respondent [FE]	0.00973
Mean Lower Court Direction Respondent 10 [FE]	0.00900
TOTAL	**0.07946**

Current Supreme Court Trends	
Mean Agreement Level of Current Court [FE]	0.00955
Std. Dev. of Agreement Level of Current Court [FE]	0.00936
Mean Current Court Direction Circuit Origin [FE]	0.00789
Std. Dev. Current Court Direction Circuit Origin [FE]	0.00850
Mean Current Court Direction Circuit Source [FE]	0.00945
Std. Dev. Current Court Direction Circuit Source [FE]	0.01021
Mean Current Court Direction Issue [FE]	0.01469
Z-Score Current Court Direction Issue [FE]	0.00832
Std. Dev. Current Court Direction Issue [FE]	0.01266
Mean Current Court Direction [FE]	0.00918
Std. Dev. Current Court Direction [FE]	0.00942
Mean Current Court Direction Petitioner [FE]	0.00863
Std. Dev. Current Court Direction Petitioner [FE]	0.00894
Mean Current Court Direction Respondent [FE]	0.00882
Std. Dev. Current Court Direction Respondent [FE]	0.00888
TOTAL	**0.14456**

Individual Supreme Court Justice Trends	
Mean Justice Direction [FE]	0.01248
Mean Justice Direction 10 [FE]	0.01530
Mean Justice Direction Z Score [FE]	0.00826
Mean Justice Direction Petitioner [FE]	0.00732
Mean Justice Direction Petitioner 10 [FE]	0.01027
Mean Justice Direction Respondent [FE]	0.00724
Mean Justice Direction Respondent 10 [FE]	0.01030
Mean Justice Direction for Circuit Origin [FE]	0.00792
Mean Justice Direction for Circuit Origin 10 [FE]	0.00945
Mean Justice Direction for Circuit Source [FE]	0.00891
Mean Justice Direction for Circuit Source 10 [FE]	0.00970
Mean Justice Direction by Issue [FE]	0.01881
Mean Justice Direction by Issue 10 [FE]	0.00950
Mean Justice Direction by Issue Z Score [FE]	0.00771
TOTAL	**0.14323**

Differences in Trends	
Difference Justice Court Direction [FE]	0.01210
Abs. Difference Justice Court Direction [FE]	0.00929
Difference Justice Court Direction Issue [FE]	0.01167
Abs. Difference Justice Court Direction Issue [FE]	0.00968
Z Score Difference Justice Court Direction Issue [FE]	0.01055
Difference Justice Court Direction Petitioner [FE]	0.00705
Abs. Difference Justice Court Direction Petitioner [FE]	0.00708
Difference Justice Court Direction Respondent [FE]	0.00690
Abs. Difference Justice Court Direction Respondent [FE]	0.00699
Z Score Justice Court Direction Difference [FE]	0.01280
Justice Lower Court Direction Difference [FE]	0.01922
Justice Lower Court Direction Abs. Difference [FE]	0.02494
Justice Lower Court Direction Z Score [FE]	0.01126
Z Score Justice Lower Court Direction Difference [FE]	0.00992
Agreement of Justice with Majority [FE]	0.00866
Agreement of Justice with Majority 10 [FE]	0.01483
Difference Court and Lower Ct Direction [FE]	0.01522
Abs. Difference Court and Lower Ct Direction [FE]	0.01199
Z-Score Difference Court and Lower Ct Direction [FE]	0.01217
Z-Score Abs. Difference Court and Lower Ct Direction [FE]	0.01150
TOTAL	**0.23391**

Figure 4. Final Feature Weights as of June 2013

Figure 12: Predicting the Behavior of the Supreme Court of the United States: A General Approach

d) Appendix 4

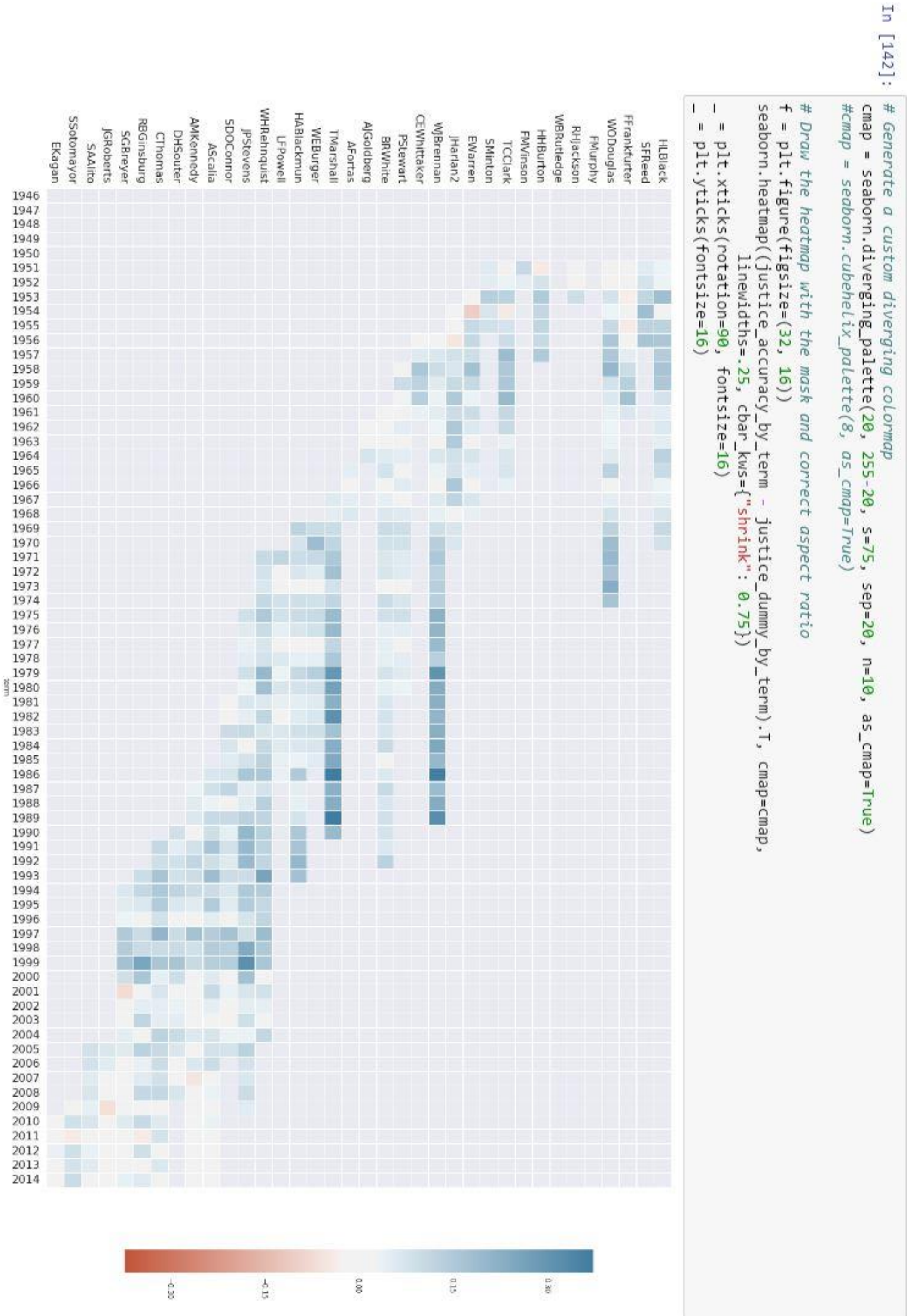

Figure 13: Heatmap produced through the prediction model by Katz et al.

e) Appendix 5

Link to the dataset with added features: https://www.dropbox.com/s/rreb9efgswdeyho/data.csv?dl=0

f) Appendix 6

(a) SVMs for the dataset with added features

```
#Read in the csv file
data<-read.csv2(file.choose(),header=TRUE)
str(data)
'data.frame':	275 obs. of  34 variables:
 $ class                              : int  1 0 0 0 0 0 0 0 0 0 ...
 $ dateDecision                       : Factor w/ 190 levels "01.03.2006","01.03.2011",..: 29
38 75 129 129 110 110 110 123 123 ...
 $ decisionType                       : int  1 1 1 1 6 1 2 2 1 1 ...
 $ majVotes                           : int  9 7 8 7 9 7 9 9 8 7 ...
 $ minVotes                           : int  0 2 0 2 0 1 0 0 0 1 ...
 $ US.supporting.Business             : int  1 1 1 0 0 0 0 0 0 0 ...
 $ Amicus.Brief.from.the.United.States      : int  1 1 1 0 0 0 1 0 0 0 ...
 $ Amicus.Brief.from.the.Chamber.of.Commerce: int  0 0 0 0 1 1 0 0 0 0 ...
 $ Amicus.Brief                       : int  1 1 1 0 1 1 1 0 1 1 ...
 $ dateArgument                       : Factor w/ 220 levels "","01.03.2006",..: 22 219 200 1
4 116 207 1 1 51 56 ...
 $ authorityDecision1                 : int  4 4 4 4 4 7 4 4 4 4 ...
 $ lawType                            : int  3 3 6 4 6 6 6 3 3 3 ...
 $ threeJudgeFdc                      : int  0 0 0 0 1 0 0 0 0 0 ...
 $ lcDisagreement                     : int  0 1 1 0 0 0 0 0 0 0 ...
 $ certReason                         : int  2 11 2 12 1 12 12 12 2 2 ...
 $ lcDisposition                      : int  2 2 5 5 9 3 9 6 3 2 ...
 $ lcDispositionDirection             : int  2 2 2 2 1 2 2 1 2 1 ...
 $ declarationUncon                   : int  1 1 1 1 1 1 1 1 1 1 ...
 $ caseDisposition                    : int  2 4 4 3 5 4 5 5 3 4 ...
 $ partyWinning                       : int  0 1 1 1 1 1 1 1 1 1 ...
 $ precedentAlteration                : int  0 0 0 0 0 0 0 0 0 0 ...
 $ decisionDirectionDissent           : int  0 0 0 0 0 0 0 0 0 0 ...
 $ issueArea                          : int  7 8 9 9 3 8 8 2 2 8 ...
 $ decisionDirection                  : int  2 1 2 1 1 1 2 2 1 1 ...
 $ caseOrigin                         : int  128 47 114 107 55 302 51 42 90 109 ...
 $ caseSource                         : int  29 28 24 8 55 300 29 31 29 23 ...
 $ issue                              : int  70040 80010 90320 90340 30140 80160 80060 20060
20400 80060 ...
 $ lawSupp                            : int  334 364 600 400 600 600 600 384 312 374 ...
 $ majOpinWriter                      : int  103 109 109 108 NA 105 NA NA 105 106 ...
 $ majOpinAssigner                    : int  111 111 111 111 111 111 111 NA 111 111 ...
 $ petitioner                         : int  158 184 113 158 207 135 24 182 161 208 ...
 $ respondent                         : int  145 125 240 158 343 138 174 151 161 412 ...
 $ term                               : int  2005 2005 2005 2005 2005 2005 2005 2005 2005 200
5 ...
 $ naturalCourt                       : int  1701 1701 1701 1701 1701 1702 1702 1702 1702 170
2 ...


#install randomForest package
install.packages("randomForest")
library("randomForest", lib.loc="~/R/win-library/3.2")

#install e1071 package
install.packages("e1071")
library("e1071", lib.loc="~/R/win-library/3.2")

#convert target variable into a factor
data$class<-as.factor(data$class)
```

```
#Formatting Dates
data$dateDecision<-as.POSIXct(data$dateDecision)
data<-na.omit(data)
data$dateArgument<-as.POSIXct(data$dateArgument)
data$datedifference<-as.numeric(difftime(data$dateDecision, data$dateArgument, units= c("days")))
data$dateDecision<-NULL
data$dateArgument<-NULL

set.seed(123)

#Reshufflelling the dataset
mydata<-data[sample(nrow(data)),]

#Creating the trainingset for 70% of the dataset
datatraining<-mydata[1:170,]

#Creating the testingset for 30% of the dataset
datatest<-mydata[171:244,]

#Use training data tune SVM
tunedmodelSVMdata <- tune.svm(class~., data = datatraining, gamma = 2^(-20:20), cost = 10^(20:20),scale =FALSE)

#Report the best value of the parameters of the tuning process
tunedmodelSVMdata$best.parameters[[1]]
[1] 0.015625
tunedmodelSVMdata$best.parameters[[2]]
[1] 1e+20
best.svm(x = class ~ ., data = datatraining, gamma = 2^(-20:20), cost = 10^(20:20), scale=FALSE)

Call:
best.svm(x = class ~ ., data = datatraining, gamma = 2^(-20:20), cost = 10^(20:20), scale = FALSE)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1e+20
      gamma:  0.015625

Number of Support Vectors:  170

finalmodelSVMdata <- svm(class~.,data=datatraining,gamma=tunedmodelSVMdata$best.parameters[[1]],cost=tunedmodelSVMdata$best.parameters[[2]],scale = FALSE)
summary<-finalmodelSVMdata

#Report the accuracy in testing data
predictionSVMdata <- predict(finalmodelSVMdata,datatest[,-1])
predictiontableSVMdata<-table(pred=predictionSVMdata,datatest[,1])
accuracySVMdata<-sum(diag(predictiontableSVMdata))/nrow(datatest)
summary(predictionSVMdata)
 0  1
74  0

summary(predictiontableSVMdata)
Number of cases in table: 74
```

```
Number of factors: 2
Test for independence of all factors:
         Chisq = NaN, df = 1, p-value = NA
         Chi-squared approximation may be incorrect
```

summary(accuracySVMdata)

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 0.7703  0.7703  0.7703  0.7703  0.7703  0.7703
```

(b) SVMs for dataset without added features

```
#Creating a dataset without the added features
modified<-data
modified$US.supporting.Business<-NULL
modified$Amicus.Brief.from.the.United.States<-NULL
modified$Amicus.Brief.from.the.Chamber.of.Commerce<-NULL
modified$Amicus.Brief<-NULL

set.seed(123)

#Reshufflelling the dataset
mymodified<-modified[sample(nrow(modified)),]

#Creating the trainingset for 70% of the dataset
modifiedtraining<-mymodified[1:170,]

#Creating the testingset for 30% of the dataset
modifiedtest<-mymodified[171:244,]

dim(modifiedtraining)
[1] 170  28
dim(modifiedtest)
[1] 74 28

#Use training data tune SVM
tunedmodelSVMmodified <- tune.svm(class~., data = modifiedtraining, gamma = 2^(-20:20), cost = 10^(20:20),scale=FALSE)
#Report the best value of the parameters of the tuning process
tunedmodelSVMmodified$best.parameters[[1]]
[1] 0.015625

tunedmodelSVMmodified$best.parameters[[2]]
[1] 1e+20
best.svm(x = class ~ ., data = modifiedtraining, gamma = 2^(-20:20), cost = 10^(20:20), scale=FALSE)

Call:
best.svm(x = class ~ ., data = modifiedtraining, gamma = 2^(-20:20), cost = 10^(20:20), scale = FAL
SE)


Parameters:
   SVM-Type:  C-classification 
 SVM-Kernel:  radial 
       cost:  1e+20 
      gamma:  0.015625 

Number of Support Vectors:  170
```

```
#Use the parameters to build the model
finalmodelSVMmodified <- svm(class~.,data=modifiedtraining,gamma=tunedmodelSVMmodified$best.parameters[[1]],cost=tunedmodelSVMmodified$best.parameters[[2]], scale=FALSE)
summary<-finalmodelSVMmodified

#Report the accuracy in testing data
predictionSVMmodified <- predict(finalmodelSVMmodified,modifiedtest[,-1])
predictiontableSVMmodified<-table(pred=predictionSVMmodified,modifiedtest[,1])
accuracySVMmodified<-sum(diag(predictiontableSVMmodified))/nrow(modifiedtest)

summary(predictionSVMmodified)
 0  1 
74  0 
summary(predictiontableSVMmodified)
Number of cases in table: 74 
Number of factors: 2 
Test for independence of all factors:
	Chisq = NaN, df = 1, p-value = NA
	Chi-squared approximation may be incorrect
summary(accuracySVMmodified)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 0.7703  0.7703  0.7703  0.7703  0.7703  0.7703 
```

(c) RF for dataset with added features

```
set.seed(123)
data.rf<-randomForest(class~.,data=data, ntree=2000)
print(data.rf)

Call:
 randomForest(formula = class ~ ., data = data, ntree = 2000) 
               Type of random forest: classification
                     Number of trees: 2000
No. of variables tried at each split: 5

        OOB estimate of  error rate: 1.23%
Confusion matrix:
    0  1 class.error
0 175  2  0.01129944
1   1 66  0.01492537

importance(data.rf)
```

```
> importance(data.rf)
                                      MeanDecreaseGini
decisionType                               0.018485690
majVotes                                   0.625997559
minVotes                                   0.610316889
US.supporting.Business                     0.215854820
Amicus.Brief.from.the.United.States        0.494462753
Amicus.Brief.from.the.Chamber.of.Commerce  0.365241684
Amicus.Brief                               0.050414490
authorityDecision1                         0.724829280
lawType                                    0.581324897
threeJudgeFdc                              0.148305843
lcDisagreement                             0.391352120
certReason                                 1.128364558
lcDisposition                              1.018262151
lcDispositionDirection                     0.815217756
declarationUncon                           0.431367776
caseDisposition                           41.102075429
partyWinning                              32.520200335
precedentAlteration                        0.006848884
decisionDirectionDissent                   0.002732698
issueArea                                  0.961501918
decisionDirection                          0.652507034
caseOrigin                                 1.699852511
caseSource                                 1.355035604
issue                                      1.927395073
lawSupp                                    1.319088683
majOpinWriter                              1.387561916
majOpinAssigner                            0.606572433
petitioner                                 1.850343657
respondent                                 2.206322613
term                                       1.138803844
naturalCourt                               0.528517370
```

(d) RF for dataset without added features

```
set.seed(123)
modified.rf<-randomForest(class~.,data=modified, ntree=2000)
print(modified.rf)
```

```
Call:
 randomForest(formula = class ~ ., data = modified, ntree = 2000)
               Type of random forest: classification
                     Number of trees: 2000
No. of variables tried at each split: 5

        OOB estimate of  error rate: 1.23%
Confusion matrix:
    0  1 class.error
0 175  2  0.01129944
1   1 66  0.01492537
```

```
importance((modified.rf))
```

```
> importance((modified.rf))
                         MeanDecreaseGini
decisionType                  0.017598118
majVotes                      0.562413452
minVotes                      0.483745144
authorityDecision1            0.684796980
lawType                       0.551443272
threeJudgeFdc                 0.149826868
lcDisagreement                0.375933494
certReason                    1.055556342
lcDisposition                 0.880762764
lcDispositionDirection        0.783208932
declarationUncon              0.363632973
caseDisposition              44.215151520
partyWinning                 32.682624634
precedentAlteration           0.002451393
decisionDirectionDissent      0.002748419
issueArea                     0.824273784
decisionDirection             0.560034782
caseOrigin                    1.549061294
caseSource                    1.250893083
issue                         1.692678370
lawSupp                       1.242754455
majOpinWriter                 1.164158895
majOpinAssigner               0.589281344
petitioner                    1.661603230
respondent                    2.122741203
term                          0.945717129
naturalCourt                  0.451459099
```

g) Appendix 7

Link to the original algorithm by Katz et al.: https://github.com/mjbommar/scotus-predict