Ministry of Education and Science of the Republic of Kazakhstan
Suleyman Demirel University

UDC 50.09

On manuscript rights

Dauylov Sanzhar

Introduce predictive analytics using the Next Best Action (NBA) models into the banking system

THESIS

Presented in Partial Fulfillment for the
Degree of Master of Science in Computing Systems and Software
(degree code: 6M070400)
Department of Computer Sciences
Faculty of Engineering and Natural Sciences

Supervisor: **Bogdanchikov Andrei**

Kaskelen, 2020

Abstract

NBA - is an approach in which each client is initially considered purely individual. It has a close correlation with Predictive analysis. Predictive or prognostic analytics is a set of techniques and methods for analyzing data to build a forecast of future events. The banking system is currently using the method to obtained certain business results from its customers and has increased loyalty, increased income, found new growth points, etc. The classical model of marketing was rather different, it repelled from its existing product line and its parameters. But new models repels from customer's inclination to purchase a particular product. The aim of this project is to investigate the field of deposit accounts of the banking system by using NBA approach and to show the benefits and possible outcomes. This approach was tested on various aspects of the banking system and showed a number of solutions which can predict the probability of a customer to create a term deposit account.

Аңдатпа

NBA - бұл әр клиент бастапқыда жеке тұлға болып саналатын тәсіл. Болжамдық талдаумен тығыз байланыс бар. Болжалды немесе болжамдық аналитика - бұл болашақ оқиғалардың болжамын құру үшін мәліметтерді талдау әдістері мен әдістерінің жиынтығы. Қазіргі уақытта банк жүйесі өз клиенттерінен белгілі бір бизнес нәтижелерін алу әдісін қолданады және адалдықты арттырды, кірісті ұлғайтты, жаңа өсу нүктелерін тапты және т.с. маркетингтің классикалық моделі мүлдем өзгеше болды, ол өзінің қолданыстағы өнімі мен параметрлерінен өзгеше болды . Бірақ жаңа модельдер тұтынушының белгілі бір өнімді сатып алуға деген бейімділігінен арылтады. Бұл жобаның мақсаты - НБА тәсілін қолдана отырып, банктік жүйенің депозиттік шоттары саласын зерттеу және оның артықшылықтары мен мүмкін нәтижелерін көрсету. Бұл тәсіл банктік жүйенің әртүрлі аспектілері бойынша сыналды және клиенттің мерзімді депозиттік шотты құру ықтималдығын болжайтын бірқатар шешімдерді көрсетті.

Аннотация

NBA - это подход, при котором каждый клиент изначально считается сугубо индивидуальным. Это имеет тесную связь с прогнозным анализом. Прогнозирующая или прогностическая аналитика - это набор методов и методов анализа данных для составления прогноза будущих событий. Банковская система в настоящее время использует метод для получения определенных бизнес-результатов от своих клиентов и имеет повышенную лояльность, увеличенный доход, найденные новые точки роста и т. Д. Классическая модель маркетинга была несколько иной, она отталкивалась от существующей линейки продуктов и ее параметров. , Но новые модели отталкивают от склонности покупателя покупать конкретный товар. Целью данного проекта является исследование области депозитных счетов банковской системы с использованием подхода НБА и демонстрация преимуществ и возможных результатов. Этот подход был опробован в различных аспектах банковской системы и показал ряд решений, которые могут предсказать вероятность того, что клиент создаст срочный депозитный счет.

Acknowledgements

Thanks to my thesis supervisors for support and useful discussion, thanks external reviewer for very useful feedback that helped to significantly improve the current work. Thanks to my family and friends.

Table of contents

1. Introduction

The theme of emotion and customer experience has been worrying bankers for quite some time. The tightening of regulations and requirements prompted banks to look for new sources of competitive advantage, in the US for example, 92% of millennials claim to have a lack of confidence in the traditional banking system and increasingly use new services. Customer experience is an indirect characteristic that translates into financial results of a bank through customer loyalty, a desire to continue to serve all of them at the same bank even in conditions of high consumer awareness and low switching costs to other financial companies. 78% of bank managers around the world surveyed by the IBM Institute for Business Value argue that customer engagement and understanding customer's needs are key conditions for the best customer experience and successful creation of new products and services. This encouraged me to take this topic, as the banks have a huge amount of data that have been accumulating for decades, which carries a great potential value. The beginning for me was a study on this topic in various sources ranging from free resources to turnkey solutions to large corporations. Each study helped me identify the pros and cons of their solutions, the importance of the quality of data attributes, and so on. Initially, it was supposed to use the data of the Kazakhstan data, but we could not agree and then found small data in the public domain. It has more than 15 attributes, an attribute was used as a prediction that indicates whether the client will open a deposit in various cases. And for implementation, one of the most popular solutions in machine learning was taken. These classifiers are widely used in various problems in life. Using the above data, I applied eight different classifiers to them and compared between them, also during work there were various problems associated with retraining, which I managed through the use of technologies such as Stratified Sampling and others. Upon completion, it was also clearly shown which classifier better in our case. In the form of continuation and improvement of work, I see with increasing data and their quality, as well as try to apply and test various parameters.

1.1 Motivation

Automated information processing systems are systems or complexes of systems designed to automate the processing of a certain type of information. In this test, the automation of the banking system is considered. For timely and high-quality processing of the ever-increasing volumes of information received by banks, the use of more and more sophisticated hardware and software is required. The banking system is not only banks, but also credit institutions, as well as specialized organizations that do not carry out banking operations, but support the activities of banks and credit institutions (cash settlement centers and clearing centers, credit shops, bank audit firms ...). For more efficient operation of banking systems, automated banking systems (ABS) have been created.

The goals of creating an ABS are:

- Reducing time for operations and paperwork, increasing the throughput of the bank.
- Decrease in the number of personnel engaged in unskilled routine work.
- Improving the quality of customer service.
- Improving the qualifications of banking personnel.
- Integration into unified banking systems.

On the domestic market, ABS classes have formed, each of which has specific consumers, from start-up banks that carry out only a limited range of ruble transactions, to leading banks that have reached a foreign level of volumes and variety of services. ABS contains a set of functions necessary for the consumer. The development of computer technology and information technology has allowed most banks to create their own computer systems, on the basis of which the main areas of banking activity were automated. The deepening of the process of automating the functioning of banking and other financial structures is accompanied by an improvement in the technology of banking operations and an increase in their manageability. Modern information technologies make it possible to coordinate the activities of bank divisions, expand interbank relations, and comprehensively solve the problems of banking analysis. Automation of information and other technologies of the bank contribute to improving the quality of service by creating automated workstations (AWS) for specialists of all levels. In the automation of banking technologies, both simple software products that allow you to fill out only a few output forms for reporting, and fairly intelligent systems that solve the problems of bank management find their place.

1.2 Aims and Objectives

The goal of my work is to investigate the influence of attributes on the result, applying various tools from machine learning, using the NBA approach. To achieve my task, I need to complete the following steps: 1. Collect a large amount of data that is suitable for my research. 2. Select several classifiers that I will use, then compare them with each other to determine the best. 3. To do data analysis, according to significant attributes and describe how this or that property influenced the decision.

1.3 Thesis Outline

The first chapter is Introduction chapter. It is this one that you are currently reading. It gives insight into the work done, and brief introduction of how to process of implementation was look like. In Chapter 2 we see how banking system work overall, review related work, some out of the box ready solutions and formulate the problem to solve. Also there was comparisons between this solutions with their pros and cons. Chapter ?? is describing the data that was used, classifications and their comparisons, and the solution to the problem. And in Conclusion chapter we conclude our conclusion about how the some attributes and how some of the influenced most, and future plan work. SDU will stand for Suleyman Demirel University, Kaskelen.

2. Background of literature review

2.1 Banking system analysis

The banking system is a combination of banks, non-banking institutions, banking infrastructure, which are in close interaction with each other and ensuring its sustainable development.

Banking systems can be classified according to various criteria.

So, depending on the type of banking relations in society, it is customary to distinguish between banking systems of a distribution, transitional and market type. A market type system is characterized by competition and regulation. The distribution type is characterized by the complete absence of market elements, strict regulation and centralization of management from a single economic center. The transitional type includes features of both the market (competition and regulation), and the distribution economic system - tight administration for some positions.

Moreover, taking into account the type of the banking system, whether distribution or market, the levels of the banking system are formed and function. In practice, one-, two- and three-level systems are encountered.

Classifying banking systems by models, we can distinguish competitive, oligopoly and monopoly models of construction. By classes - national, supranational and world banking systems.

The level of specialization distinguishes between universal and specialized banking systems. With a specialized model, it is forbidden to combine credit and investment activities. The universal model of the banking system, on the contrary, allows a combination of lending and investment.

Depending on the degree of development of the banking system, in practice there are:

- extensive model. It is characterized by a limited number of banking services, an aggressive policy on the market of assets and liabilities, a low degree of diversification, a high concentration of risks, a low level of development of competition and market discipline;

- intensive model. It is distinguished by a high level of development of competition, a high degree of transparency and market discipline, the presence of a branched modern infrastructure, a high degree of capitalization of banks, balanced business conduct and sustainability, and the reliability of information published and provided to the control and supervision bodies.

The analysis is a set of special knowledge that is necessary to study the activities of a commercial bank.

Functional analysis of the banking system:

- a full assessment of the results of the bank and the financial condition of the organization.
- determination of the shortcomings of banking management.
- obtaining a quantitative assessment of the economic potential of the bank.
- development of a strategy for the further development of the bank, based on the prevailing conditions in the financial markets.

Principles of banking system analysis

- Define the basics of organization and analysis.
- Disclose the content of the requirements presented by users to the results obtained during the analysis.

The principles of banking analysis include the following:

1. The principle of external interaction. When analyzing, it is necessary to take into account the economic, social and political structure of the country where the bank operates

2. The principle of a scientific approach. The basis of the analysis should be a comprehensive scientific approach. It is also necessary to take into account new technical developments in the field of banking activity assessment;

3. The principle of complexity. When conducting an analysis, it is necessary to take into account all aspects of the activities of a credit institution in order to develop further comprehensive measures to eliminate deficiencies and increase efficiency.

4. The principle of consistency. Each element of the financial activities of a commercial bank should be considered as part of a single system, this determines the systematic analysis.

5. The principle of accuracy. Accuracy in the analysis plays a huge role, it must be supported by appropriate calculations, reports and conclusions.

6. The principle of practical significance. The results of the analysis should be actively applied in the future activities of a commercial bank.

7. The principle of planning. The regularity of the analysis is its main quality. Each commercial bank should develop a plan for the analysis of its activities;

8. The principle of efficiency. Efficiency is an important quality of the analysis, since its results correspond to exactly the current market conditions;

9. The principle of objectivity. The analysis should be carried out by employees with special qualifications, permission to conduct it, as well as experience in conducting the analysis.

10. The principle of comparability. The results of the analysis should be comparable with similar data obtained from other banks.

11. The principle of profitability. This principle characterizes the minimum cost of analysis.

Banking System Analysis Functions

1. Justification of the plans of the banking system;

2. Monitoring the implementation of plans;

3. Search for ways to increase efficiency;

4. Definition of measures for the use of additional resources identified in the analysis;

5. Assessment of the final results of the bank;

6. The study of the impact of the legislative framework on the financial activities of the bank.

Analysis Steps

1. Preparation for analysis. At this stage, the participants in the analysis are determined.

2. The preliminary stage. At this stage, the following actions are performed:

- Verification of collected data.
- Generalization of data.
- Calculation of standards for absolute and relative indicators.

3. The analytical stage. This stage includes: obtaining the characteristics of the necessary indicators, the ratio of the indicators of the analyzed bank with the indicators of competing banks, the final conclusion.

4. The final stage. At the end of the analysis process, an assessment is made of the level and quality of management services, the creation of forecasts and recommendations.

Banking System Analysis Classification

1. Depending on the spectrum of questions studied:

- Thematic.
- Full

2. The scope of the research objects:

- Continuous.
- Selective.

3. Depending on the purpose of the analysis:

- initial.
- operational.
- final.
- promising.

4. Depending on the object of study:

- functional.
- structural.
- operational cost

2.2 Ready solutions

Microsoft offers Veripark's [1] Next Best Action (N.B.A.) is a leading customer-centric technology that considers all the possible actions during a customer interaction and recommends the next best one thereby increasing the likelihood of positive response. For some, budgetary establishments, advances are set at the focal point of the associations as they give a key fixing in deciding authoritative productivity. Money related organizations are progressively aware of improving advance beginning and adjusting procedures to expand productivity and choice speed. They additionally intend to improve their abilities to give unrivaled client experience.

Overseeing intricate, adaptable and multiphase loaning forms requires exceeding expectations in activities. VeriPark's advance beginning and adjusting arrangement VeriLoan created on Microsoft Dynamics 365 offers the top tier start to finish answer for monetary establishments. It gives a full scope of functionalities that smoothes out credit beginning and adjusting forms, while augmenting effectiveness and diminishing danger.

VeriLoan is an advance start arrangement that encourages money related establishments to. Kill manual preparing via robotizing the paper-based procedures, Empower monetary establishment representatives to invest energy in things that issue, for example, deals exercises and client relationship the executives, rather than organization, Upgrade capacities, Offer customized process streams, strategies and rules. One application to control all start channels.

Catching the credit demand is the first and most significant piece of an advance start arrangement. VeriLoan assists with catching the information and oversee consistent, steady and connecting with client travels on a similar stage, regardless of what channel or blend of channels you decide to utilize. Numerous directs being overseen in a similar stage makes far reaching advance beginning procedures that empower a comprehensive survey of like Advanced Loan Origination, Moment Loan Origination, Portable Loan, Origination and Trader based Loan Origination.

VeriLoan's work process innovation to oversee work steps in the advance beginning and adjusting forms diminishes the measure of excess. It incorporates rule motor based scoring and all the functionalities and structures that credit processors require while handling an advance. It gives broad parameterization, reusable business rules and simple access to all structures, checks, loan specialists and outsiders.

RapidMiner [2] is a software platform for analytics teams that unites data prep, machine learning, and predictive model deployment, also one of the solutions is Next Best Action. Their slogan is "One platform, does everything", they mean visual workflow, team collaboration, model management, deployment and work with technologies like Hadoop and Spark. For example clients consider their to be with you as balanced, and they expect that you do, as well. As interesting people, they would prefer not to feel they're getting impacted with promoting messages focused on thousands or a large number of others. Getting proceeded with reliability requires doing precisely what's directly for every client at every second, be it making an offer, introducing helpful substance, or giving unique client support. Information science utilizes life-occasion designs, purchasing conduct, online networking cooperations, and different bits of knowledge to choose which moves ought to be made for every client increment dedication, strengthen collaboration with your association and drive incomes.

Cloud Sense [3] platform has a good interface, tutorials and before using it, you can see the demo version. For an organization that uses Salesforce, a most popular CRM system, CloudSense easily can be integrated. The abyss between the experience your clients request and what you right now convey is developing. However, your kin can't convey on cutting edge client desires alone. To process, comprehend and saddle the sheer volume of information that is important to convey customized encounters and unrivaled business results your kin need help. Virtuoso tackles Artificial Intelligence (AI) and NBA to insightfully direct your kin, enabling them to settle on decisions that charm clients and concentrate most extreme worth. Make a greater amount of your client information on Salesforce – influence prescient knowledge, versatile examination and savvy business rules to convey extraordinary encounters and business results progressively with Genius.

Pega [4] they're a software solution called Pegasystems. Good company with a couple of success cases. In the platform, they have Customer Decision Hub, Intelligent Guided Selling, Contextual Next-Best-Action Marketing Intelligent Guidance for Customer Service.A portion of your best clients are pondering leaving you. They aren't getting the customized administration that the present shoppers request and there are a lot of banks out there to look over. Without a doubt, you may make a last-dump offer of deferred expenses or a free record, yet it may not be sufficient to shield them from exiting the entryway. Imagine a scenario in which you could foresee what your clients require and give it when and how they need it. Rather than group and impact, you'll have the ability to convey the correct proposal to the ideal individual at the ideal time. With Pega's demonstrated AI innovation, you can customize any maintenance offer, on any channel. You can proactively start a deals or maintenance discussion with a client the second an issue emerges. Far better, you can proactively get in touch with them before they experience a terrible encounter. Foreseeing what's to come. Boosting dedication. What's more, eclipsing the opposition. Not an issue with Pega.

NGData [5] offers the solution of Next Best Action as a whole marketing system. This decision is an excellent solution for the organization that works with clients and they need a CRM System. Today, purchasers expect increasingly pertinent, associated, and opportune commitment from the organizations they trust, conveyed all the time and lined up with their life objectives and yearnings. With NGDATA's out-of-the-case arrangements you can increase constant bits of knowledge into buyers while consequently driving important encounters that lead to trust and brand promotion. Our industry out-of-the-case answers for budgetary administrations, cordiality, telcom, media and amusement, utilities, and retail are based on NGDATA's Intelligent Engagement Platform and empower you to promptly make an incentive at each phase of the client lifecycle by giving 1:1 omni-channel and constant commitment. Client DNA, ongoing crowds, offers, and encounters are accessible pre-bundled so you can rapidly begin improving client securing, strategically pitch/upsell, enactment, overhauling, maintenance, and backing. Simultaneously, you have the adaptability to refine and include custom use-cases, constantly rethinking the eventual fate of client driven commitment.

Jacada [6] prefer a flexible system for working with their platform. Also, they have a nice intelligent chatbot. Their Agent Desktop is prime real-estate in the bid to win and keep customers. Specialists are working in progressively complex situations. Call types are getting progressively mind boggling (all things considered, the more straightforward calls are being dealt with without anyone else administration arrangements), the administrative scene is getting all the more testing, and organizations are offering more items and administrations. The present call place specialist is encircled by innovation. Specialists are regularly left to using obsolete static contents, or more regrettable, depending on ineffectively remembered contents from preparing directed months prior. Specialists likewise need to keep an eye out for administrative, process, or lawful prerequisites. The outcome? Long call times and baffled clients (and specialists), a high mistake rate seriously affecting first call goals, and, maybe more worryingly, activities outside of affirmed organization or legitimate administrative prerequisites.

Also, early in this year, there was a good article [7], about how Sberbank works with BigData and what kind of tools they are using. Their tools and methods can be as good for a start. They have a competence center for big data. Accordingly, they are engaged in Hadoop, everything that is near it, and everything that allows us to process big data. In addition to Hadoop, these are key-value repositories, distributed text indexes, auxiliary coordinators of Zookeeper, MPP solutions from Hadoop's orbit, like Impala (if it can be called full-fledged MPP - but, nevertheless, it claims to be). This, of course, Spark, Kafka and other popular things. It is technologically. In theory, Hadoop for a bank is a platform that allows you to process large amounts of internal and external banking data. At conferences, they often talk about SBS, and this is a transactional system. Hadoop is an analytical system. PPRB processes more up-to-date information, and Hadoop - accumulates the entire historical volume of information. This is transaction data, and some kind of client data, accounts, operations. This applies to individuals and legal entities. There is a number of external data, for example, Spark Interfax for corporate use, some sources of credit history, and so on.

All this data is aggregated in Hadoop to build customer profiles, make decisions for retailers about marketing campaigns, and calculate the likelihood of an outflow of customers. To intelligently build different schedules, for example, work schedules of units. Recently, they built a model that allows us to identify failures in Sberbank's mobile application based on reviews on Google Play — that is, they began to analyze arbitrary text. They experimented with image analysis - this is necessary in order to better determine the bank's client by video or by photo from cameras. The bank has many of these areas.In addition, there is infrastructure hardcore when we need to integrate with other banking systems. When you need to download a huge amount of data in realtime mode in WAY4, SmartVista or CEN. That is, this is some combination of Big Data and high load. This raises interesting purely technological issues. If usually at all conferences they say that Big Data and machine learning are one and the same thing, then here they are not. On the way to building an enterprise platform, they realized that there are many purely technical problems that out of the box are not solved in the Big Data world. In particular, it is a simple task to get a stream of real data on Hadoop. But on streams the size of only a few terabytes per hour, it from a simple turns into a separate big task. Providing realtime access to data on Hadoop is another area that is not very covered by turnkey solutions.

2.3 Other solutions

One of the companies in New Zealand used a lot of data for improving credit scoring models [8]. They construct a baseline model based solely on the existing scoring features obtained from the loan application form, and a second baseline model based solely on the new bank statement-derived features. A combined feature model is then created by augmenting the application form features with the new bank statement derived features. Right now, explore the degree to which highlights got from bank proclamations gave by advance candidates, and which are not proclaimed on an application structure, can upgrade a credit scoring model for a New Zealand loaning organization. Investigating the capability of such data to improve credit scoring models right now not been concentrated beforehand. They have develop a benchmark model dependent on the current scoring highlights acquired from the advance application structure, and a subsequent pattern model based exclusively on the new bank explanation inferred highlights. A joined element model is then made by enlarging the application structure highlights with the new bank articulation inferred highlights.

Our exploratory outcomes show that a joined component model performs superior to both of the two pattern models, and that some of the bank articulation determined highlights have esteem in improving the credit scoring model. As is frequently the situation in credit scoring, our objective information was profoundly imbalanced, and Naive Bayes was seen as the best performing classifier, outflanking various different classifiers regularly utilized in credit scoring. Future experimentation with Naive Bayes on other profoundly imbalanced credit scoring informational collections will assist with affirming regardless of whether the classifier ought to be all the more regularly utilized in the credit scoring setting. The oddity of our commitment is in the examination of getting highlights from bank proclamation information gave by advance candidates, which are not pronounced on the advance application structure, to be utilized in credit scoring models. As far as we could possibly know, this has not been concentrated already. For this examination, right off the bat we build two gauge models. The principal, in view of on the current scoring highlights, and the second, in light of on the new bank proclamation determined highlights. Also, we make a joined component model by consolidating the current scoring highlights with the new bank articulation determined highlights.

Data can be used also like how Monetary establishments get gigantic measures of information about client exchanges and cash moves, which can be considered as a huge diagram powerfully changing in time. In this work [9], They center around the assignment of anticipating new connections in the system of bank customers and treat it as a connection expectation issue. They propose another chart neural system model, which utilizes the topological structure of the system as well as rich time-arrangement information accessible for the diagram hubs and edges. They assess the created strategy utilizing the information gave by a huge European bank for quite a long while. The proposed model outflanks the current methodologies, counting other neural system models, with a noteworthy hole in ROC AUC score on interface forecast issue and furthermore permits to improve the nature of credit scoring. The considered dataset is acquired from one of the huge European banks. The information comprises of client exchanges and cash moves between clients during five years. All the information is depersonalized with every exchange being portrayed by timestamp, sum and money. Therefore, they watch a chart G(V, E) with a lot of vertices V and a lot of edges E. Here, an edge (I, j) E implies that there was at any rate one exchange between a couple of customers I and j over the watched timespan. Every hub I V is spoken to by a period arrangement of exchanges for customer I, while each edge (I, j) E is spoken to by a period arrangement of moves between customers I and j. At last, they acquire a gigantic 86-million hubs chart with around 4 billions of edges.

In this article [10] we can find interesting solutions that solves two problems of clients of the Hungarian bank otpbank. The test was separated into two undertakings. The primary errand was to anticipate for each bank office the number of visits for a lot of clients, the subsequent errand was to anticipate, regardless of whether a client will apply for a charge card in the following a half year. For these assignments, anonymized client data (for example age, area, salary, sex) and bank exercises (for example what has been purchased, where and when) were given. A named informational index for 2014 was caused accessible which to can be utilized for administered machine figuring out how to foresee the objectives for a disjoint arrangement of clients for 2015. The assessment measure for Task 2 is the territory under the ROC bend (AUC), a very basic measure for imbalanced characterization issues.

The assessment measure for Task 1 is somewhat increasingly colorful. Basic data of the clients was accessible including age, area, pay and sexual orientation. While sexual orientation is essentially double, different highlights were as of now binned into three classifications. We utilized this data as highlights subsequent to changing them by means of one-hot encoding. Moreover, the inner arrangement of a bank whether the client is considered as rich or not was allowed for every month. We recognized clients of following five classes: clients that have been named 1) well off in completely watched months, 2) not well off in totally watched months, 3) first rich and afterward changed to not affluent, 4) first not well off and afterward changed to rich, 5) the individuals who changed their grouping more than once. Applying one-hot encoding, we included this data as highlights. At long last, the data in what month the client has a charge card of the bank was given. Similarly to the five classifications of the well off grouping, we made classifications for the Visa time-arrangement data.

Regularization methods are broadly used to improve the consensus, strength, and productivity of profound convolutional neural systems (DCNNs). In this paper[11], they propose a novel methodology of managing DCNN convolutional pieces by an organized channel bank. Contrasting and the current regularization strategies, for example, '1 or '2 minimization of DCNN part loads and the portion symmetry, which overlook test relationships inside a bit, the utilization of channel bank in regularization of DCNNs can shape the DCNN parts to regular spatial structures and highlights (e.g., edges or surfaces of different directions and frequencies) of common pictures. Then again, not at all like legitimately making DCNN parts fixed channels, the channel bank regularization despite everything permits the opportunity of improving DCNN loads by means of profound learning. This new DCNN plan procedure intends to join the best of two universes: the incorporation of auxiliary picture priors of customary channel banks to improve the vigor and all inclusive statement of DCNN arrangements and the capacity of present day profound figuring out how to demonstrate complex non-straight capacities covered up in preparing information.

Test results on object acknowledgment assignments show that the proposed regularization approach guides DCNNs to quicker union and preferable speculation over existing regularization strategies for weight rot and portion symmetry. With the exception of being free of repetition, symmetrical parts don't think about spatial connection between's the loads inside a given part, and henceforth independent of spatial structures. A few endeavors were made to decrease the intricacy of model by including priors in the parts. Bruna and Mallat proposed a strategy called convolutional dispersing systems in which they utilized fixed fell wavelets to break down pictures [2]. In spite of the fact that the technique had great execution on explicit datasets, it diminishes the ability of CNNs. The wavelet earlier is excessively unbending to successfully describe an extraordinary assortment of obscure picture structures. Correspondingly, Chan et. al. in [4] proposed a arrange engineering called PCANet so as to make channel banks in the layers dependent on a PCA decay of input pictures. This strategy can learn convolutional bits from the sources of info, however the yield can't influence the channel bank plan. This is in strife with the plan goal of DCNNs, which is to learn convolutional portions regarding yields not simply inputs. For example, for order undertakings, the objective is to learn contingent probabilities of yield information according to the info.

2.4 Comparison of solutions

In the writings above you can see many resembling solutions that has been done as a research, but all of them decides not our problem. Because all of them has different data and goals that they wanna solve, from all that decision we can see which kind of data is more valuable for their classification models that they choose.

And also there are some ready solutions that could help somehow, but their problem is that they are limited, cost a lot, and even after buying we should adapt under our needs, which also takes not small amount of money. Some of them has high requirements, like there shouldn't be non or unnecessary instances. Most of the valuable thing here is time that you can start not from the beginning, but from some solution that can be adapted and configured, to start as soon as possible, but as we see it will cost a lot.

In this thesis I will try get all of the pros and cons, because I know from works above, how instances in my dataset can be configured to get high results.

3. Implementation of NBA approach

3.1 About Data Set

This is the classic marketing bank dataset uploaded originally in the UCI Machine Learning Repository. The dataset gives you information about a marketing campaign of a financial institution in which you will have to analyze in order to find ways to look for future strategies in order to improve future marketing campaigns for the bank.

INSTANCE NAME	CATEGORIES
age	1-99
job	admin, blue-collar, entrepreneur, housemaid, management, retired, self-employed, services, student, technician, unemployed, unknown
marital	divorced, married, single, unknown; note: 'divorced' means divorced or widowed
education	primary, secondary, tertiary and unknown
default (has credit in default?)	no, yes, unknown
housing (has housing loan?)	no, yes, unknown
loan (has personal loan?)	no, yes, unknown
balance (Balance of the individual)	

Table. 3.1: Table of Bank Client Data

- Bank client data: (see Table 3.1)
- Related with the last contact of the current campaign: (see Table 3.2)

INSTANCE NAME	CATEGORIES
contact (contact communication type)	cellular, telephone
month (last contact month of year)	jan, feb, mar, apr, may, june, july, aug, sep, oct, nov, dec
day (last contact day of the week)	mon, tue, wed, thu, fri
duration (last contact duration, in seconds)	Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed.

Table. 3.2: Related with the last contact of the current campaign

- Other attributes:(see Table 3.3)

- Output variable (desired target): 21 - y - has the client subscribed a term deposit? (binary: 'yes','no')

Summary

- Mean Age is approximately 41 years of age. (Least: 18 years of age and Maximum: 95 years of age.)

- The mean balance is 1,528. Nonetheless, the Standard Deviation (sexually transmitted disease) is a high number so we can comprehend through this that the parity is intensely dispersed over the dataset.

INSTANCE NAME	CATEGORIES
campaign (contact communication type)	number of contacts performed during this campaign and for this client
pdays (last contact month of year)	number of days that passed by after the client was last contacted from a previous campaign
previous (last contact day of the week)	number of contacts performed before this campaign and for this client
poutcome (outcome of the previous marketing campaign)	failure, nonexistent, success

Table. 3.3: Other attributes

- As the information data said it will be smarter to drop the length segment since span is profoundly associated in whether a potential customer will purchase a term store. Additionally, duration is obtained after the call is made to the potential client so if the objective customer has never gotten calls this element isn't unreasonably helpful. The motivation behind why span is exceptionally connected with opening a term store is on the grounds that the more the bank converses with an objective customer the higher the likelihood the objective customer will open a term store since a higher length implies a higher premium (duty) from the potential customer.

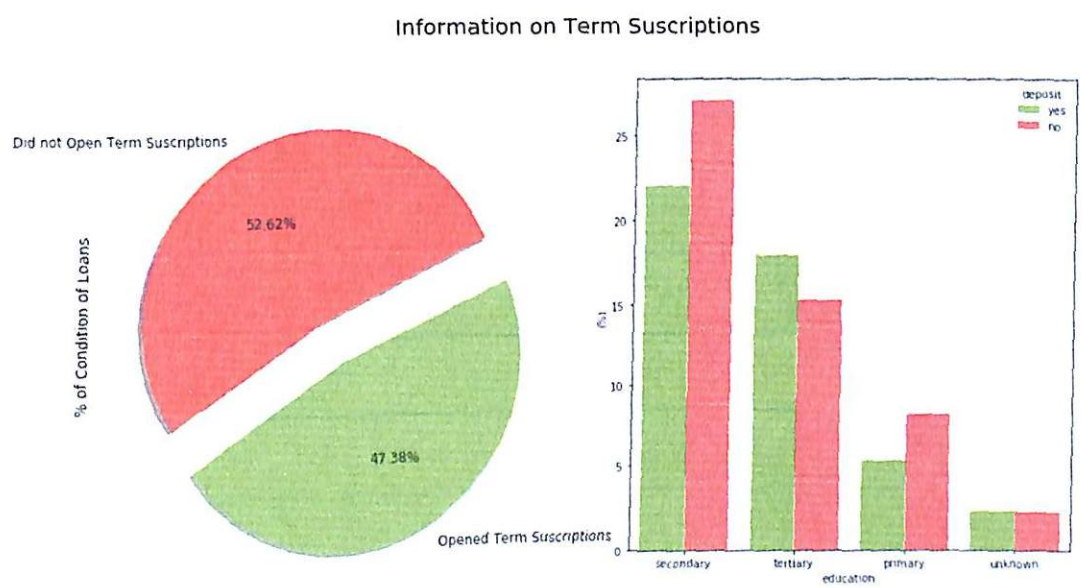

Figure. 3.1: Terms of Subscription

Luckily, there are no missing qualities. In the event that there were missing qualities we should fill them with the middle, mean or mode. I will in general utilize the middle yet right now is no compelling reason to fill any missing qualities. This will make our activity simpler!

Below you can see (Figure 3.1) deposit term subscription and his relation to education. All other graphs that not correlated and plotted as a single one, you can see Appendix A.1.

Analysis by Occupation:

- Number of Occupations: Management is the occupation that is increasingly pervasive right now.

- Age by Occupation: As expected, the resigned are the ones who have the most noteworthy middle age while understudy are the least. (Figure A.2)

- Balance by Occupation: Management and Retirees are the ones who have the most elevated equalization in their records. (Figure A.1)

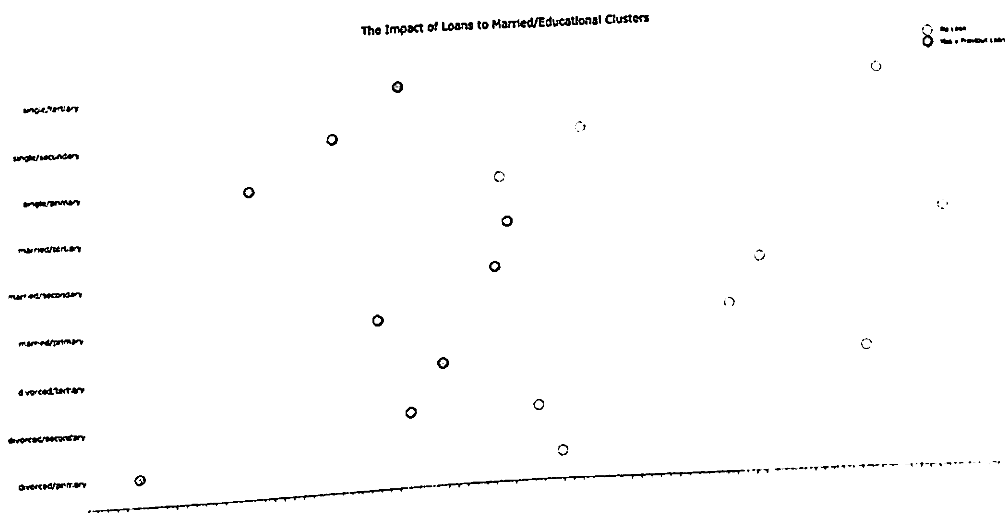

Figure. 3.2: Loans

Marital Status Well right now didn't locate any noteworthy bits of knowledge other than most separated from people are down and out. No big surprise since they need to part budgetary resources! In any case, since no further bits of knowledge have been discovered we will continue to grouping conjugal status with training status. How about we check whether we can discover different gatherings of individuals in the example populace. (see in Appendix A.9 and A.10)

Additionally, duration is obtained after the call is made to the potential client so if the objective customer has never gotten calls this element isn't unreasonably helpful. The motivation behind why span is exceptionally connected with opening a term store is on the grounds that the more the bank converses with an objective customer the higher the likelihood the objective customer will open a term store since a higher length implies a higher premium (duty) from the potential customer

Clustering Marital Status and Education:

- Marital Status: As talked about already, the effect of a separation significantly affects the parity of the person. (see Figure 3.3)

- Education: The degree of training additionally significantly affects the measure of parity a possibility has.

- Loans: Whether the possibility has a past advance significantly affects the measure of parity the person has. (see Figure 3.2)

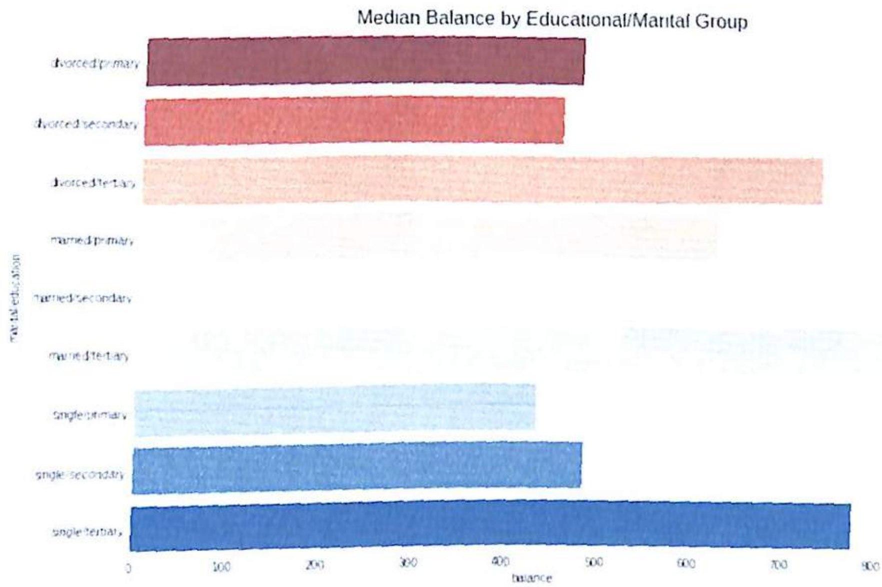

Figure. 3.3: Marital Status

Campaign Duration:

- Campaign Duration: Hmm, we see that length has a high relationship with term stores meaning the higher the span, the almost certain it is for a customer to open a term store.

- Average Campaign Duration: The normal crusade length is 374.76, how about we check whether customers that were over this normal were bound to open a term store.

- Duration Status: People who were over the length status, were bound to open a term store. 78% of the gathering that is better than expected in length opened term stores while those that were beneath normal 32%

opened term store accounts. This discloses to us that it will be a smart thought to target people who are in the better than expected classification. (see Figure 3.4)

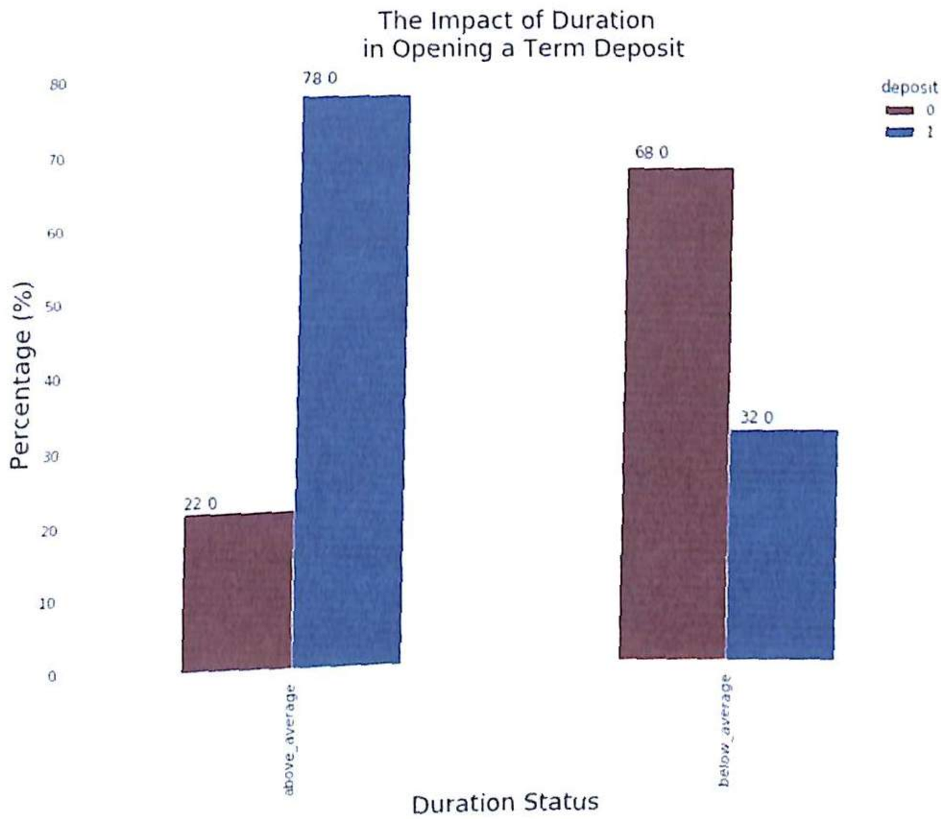

Figure. 3.4: Median Balance by Educational/Marital Group

3.2 Classification of models

Nowadays we have wonderful libraries, that we can get out of the box tools. That is why we can classify by several at once. Here I take:

Logistic Regression

Logistic regression is one of the statistical classification methods using Fisher's linear discriminant. She is also among the top frequently used algorithms in data science. In this article, the essence of logistic regression is described in such a way that it becomes clear even to people who are not very close to statistics.

The main idea of logistic regression

Unlike conventional regression, the logistic regression method does not predict the value of a numerical variable based on a sample of the initial values. Instead, the value of a function is the likelihood that a given source value belongs to a particular class. For simplicity, let's assume that we only have two classes (see Multiple Logistic Regression for tasks with a large number of classes) and the probability that we will determine is the image of the probability that some value belongs to the "+"class. And of course image. Thus, the result of logistic regression is always in the interval [0, 1].

The dependent variable in the multinomial logistic regression model can be measured in an ordinal or nominal scale. For example, mean political preferences in elections or choice of a trademark by a consumer. However, for a dependent ordinal variable, it is better to use a special ordinal regression model. Independent variables can be categorical or quantitative. Categorical independent variables are called factors. Quantitative independent variables are called covariates.

The main idea of logistic regression is that the space of initial values can be divided by a linear boundary (i.e. a straight line) into two regions corresponding to the classes. So, what is meant by a linear boundary? In the case of two measurements, it is simply a straight line without bends. In case of three - a plane, and so on. This boundary is set depending on the available input data and the training algorithm. For everything to work, the source data points must be divided by a linear border into the two above-mentioned areas. If the source data points satisfy this requirement, then they can be called linearly separable. (Look at the 3.5).

In the multinomial logistic regression model for each category of the dependent variable, the binary logistic regression equation is constructed. In this case, one of the categories of the dependent variable becomes the reference and all other categories are compared with it. The multinomial logistic regression equation predicts the probability of a dependent variable belonging to each category from the values of independent variables.

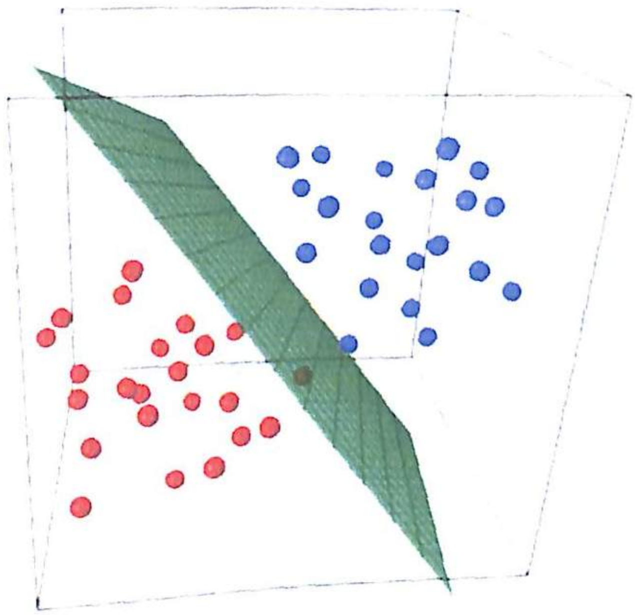

Figure. 3.5: Logistic Regressions

The indicated dividing plane is called the linear discriminant, since it is linear in terms of its function, and allows the model to separate, discriminate points into different classes.

If it is impossible to linearly separate the points in the original space, it is worth trying to convert the feature vectors into a space with a large number of dimensions, adding additional interaction effects, members of a higher degree, etc. Using a linear algorithm in such a space provides certain advantages for learning a nonlinear function, since the boundary becomes nonlinear when returning to the original space.

But how is the linear boundary used in the logistic regression method to quantify the probability of data points belonging to a particular class?

K Nearest Neighbors Classifier

kNN stands for k Nearest Neighbor or k Nearest Neighbors - this is one of the simplest classification algorithms, also sometimes used in regression tasks. Due to its simplicity, it is a good example from which you can begin your acquaintance with the field of Machine Learning. This article describes an example of writing the code for such a classifier in python, as well as visualizing the results.

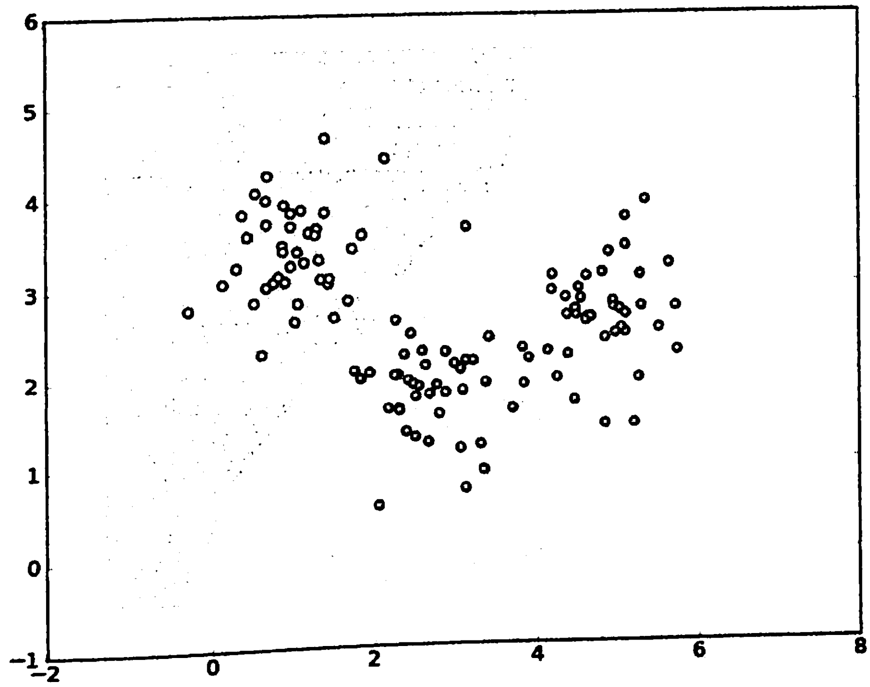

Figure. 3.6: KNN Classification example

The classification task in machine learning is the task of assigning an object to one of the predefined classes based on its formalized features. Each of the objects in this problem is represented as a vector in N-dimensional space, each dimension in which is a description of one of the features of the object. Suppose we need to classify monitors: measurements in our parameter space will be the diagonal in inches, aspect ratio, maximum resolution, HDMI interface, cost, etc. The case of classifying texts is somewhat more complicated, they usually use a term-document matrix [12].

To train the classifier, you must have a set of objects for which classes are predefined. This set is called a training set, its marking is done manually, with the involvement of specialists in the studied area. For example, in the task of Detecting Insults in Social Commentary[13] for pre-assembled tests of comments, a person puts down the opinion whether this comment is an insult to one of the participants in the discussion, the task itself is an example of binary classification. In the classification problem, there can be more than two classes (multiclass), each of the objects can belong to more than one class (intersecting).

Algorithm. To classify each of the objects of the test sample, the following operations must be performed sequentially:

- Calculate the distance to each of the objects in the training set.

- Select k objects of the training set, the distance to which is minimum.
- The class of a classified object is the class most often found among k nearest neighbors.

kNN is one of the simplest classification algorithms, so it often turns out to be ineffective in real problems. In addition to the accuracy of classification, the problem of this classifier is the speed of classification: if there are N objects in the training set, M objects in the test set, and the space dimension K, then the number of operations for classifying the test set can be estimated as O (K * M * N). Nevertheless, the kNN algorithm is a good example to get started with Machine Learning.

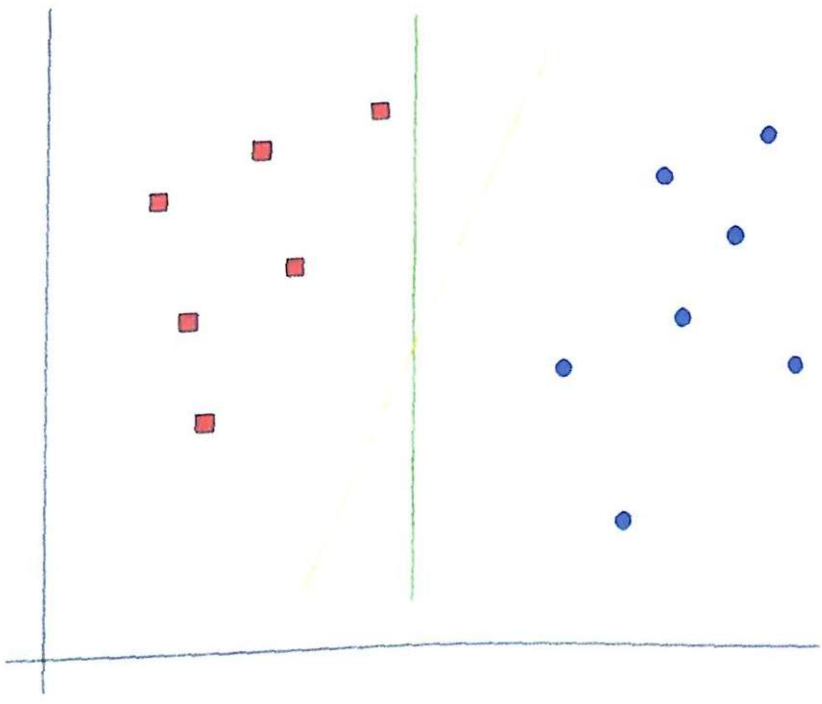

Figure. 3.7: Linear SVM

Linear SVM

The Support Vector Machines Method or SVM (from the English Support Vector Machines) is a linear algorithm used in classification and regression problems. This algorithm is widely used in practice and can solve both linear and nonlinear problems. The essence of the "Machines" of Support Vectors is simple: the algorithm creates a line or hyperplane that divides the data into classes. The main task of the algorithm is to find the most correct line or hyperplane dividing the data into two classes. SVM is an algorithm that receives data at the input and returns such a dividing line.

The main task of the algorithm is to find the most correct line or hyperplane dividing the data into two classes. SVM is an algorithm that receives data at the input and returns such a dividing line.

Consider the following example. Suppose we have a data set, and we want to classify and separate the red squares from the blue circles (let's say positive and negative). The main goal in this task will be to find the "ideal" line that will separate these two classes.(see Figure 3.7)

In the case of the green line - it is located too close to the red class. Despite the fact that she correctly classified all objects of the current data set, such a line will not be generalized - it will not behave just as well with an unfamiliar data set. The task of finding a generalized separating two classes is one of the main tasks in machine learning. Now that we have familiarized ourselves with the logic of the algorithm, we move on to the formal definition of a hyperplane. A hyperplane is an n-1 dimensional subplane in an n-dimensional Euclidean space that divides the space into two separate parts. For example, imagine that our line is represented as a one-dimensional Euclidean space (i.e. our data set lies on a straight line). Select a point on this line. This point will divide the data set, in our case the line, into two parts. The line has one measure, and the point has 0 measures. Therefore, a point is a hyperplane of a line. For the two-dimensional dataset that we met earlier, the dividing line was the same hyperplane. Simply put, for an n-dimensional space there is an n-1 dimensional hyperplane dividing this space into two parts.

In this example, there are several decision thresholds that we can define for this particular sample. Pay attention to the direct (presented on the chart as a green line) decision threshold. It is quite simple, and for this reason, several objects were classified incorrectly. These points that were classified incorrectly are called outliers in the data.

We can also adjust the parameters in such a way that in the end we get a more curved line (light blue decision threshold), which will classify absolutely all the training sample data correctly. Of course, in this case, the chances that our model will be able to generalize and show equally good results on new data are catastrophically small. Therefore, if you are trying to achieve accuracy when training the model, you should aim at something more even, direct. The higher the "C" number, the more entangled the hyperplane will be in your model, but the higher the number of correctly-classified objects in the training set. Therefore, it is important to "twist" the model parameters for a specific data set in order to avoid retraining but, at the same time, achieve high accuracy.

In the official documentation, the SciKit Learn library says that the gamma determines how far each of the elements in the data set has an influence in determining the "ideal line". The lower the gamma, the more elements, even those that are far enough from the dividing line, take part in the process of choosing this very line. If the gamma is high, then the algorithm will "rely" only on those elements that are closest to the line itself. The main task of the algorithm is to find the most correct line or hyperplane dividing the data into two classes. SVM is an algorithm that receives data at the input and returns such a dividing line.

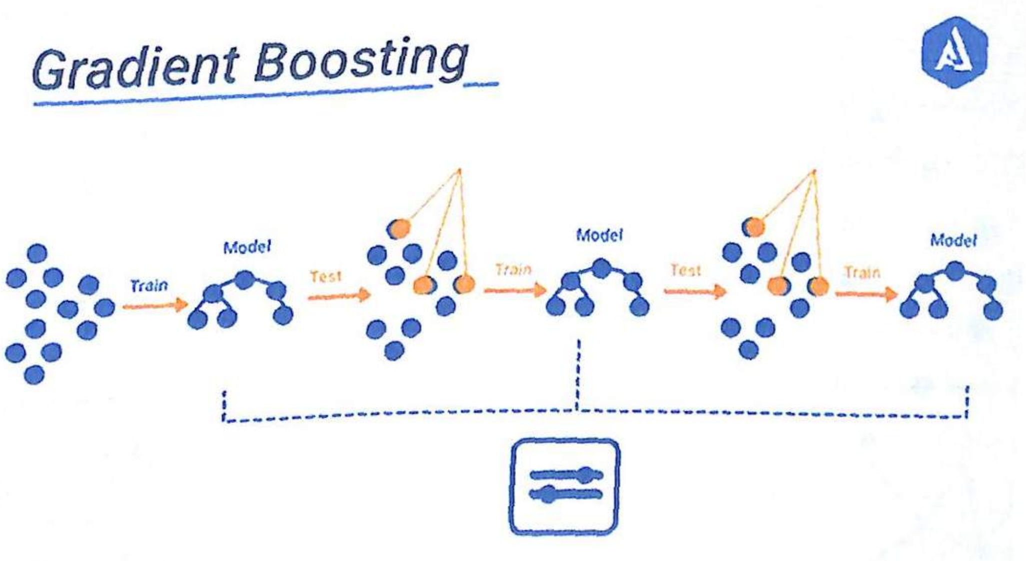

Figure. 3.8: Gradient boosting example

Gradient Boosting Model

Most people involved in data analysis have heard about boosting at least once. This algorithm is included in the everyday "gentleman's set"of models that are worth trying in the next task. Xgboost is often associated with the standard recipe for winning ML competitions, giving rise to a meme about "stack xgboosts."And boosting is an important part of most search engines, sometimes acting as their business card. Let's see how boosting came about and evolved for general development.

Boosting Story

It all started with the question of whether it is possible to get one strong one from a large number of relatively weak and simple models. By weak models, we mean not just small and simple models like decision trees, as opposed to more "strong" models, such as neural networks. In our case, weak models are arbitrary machine learning algorithms, the accuracy of which can be only slightly higher than random guessing.

An affirmative mathematical answer to this question was found quite quickly, which in itself was an important theoretical result (a rarity in ML). However, it took several years before the emergence of workable algorithms and Adaboost. Their general approach was to eagerly construct a linear combination of simple models (basic algorithms) by weighting the input data. Each subsequent model (as a rule, a decision tree) was constructed in such a way as to give greater weight and preference to previously incorrectly predicted observations.

Adaboost worked well, but due to the fact that there were few justifications for the algorithm working with its add-ons, a full range of speculations arose around them: someone considered it a super-algorithm and a magic bullet, someone was skeptical and shared the opinion that this hardly applicable approach with overfitting. This was especially true for applicability on high emission data, to which Adaboost turned out to be unstable. Fortunately, when the professors of the Stanford Department of Statistics, who had already brought the world Lasso, Elastic Net, and Random Forest, got involved, in 1999 Jerome Friedman introduced a generalization of the achievements of boosting algorithms - gradient boosting, aka Gradient Boosting (Machine), aka GBM. With this work, Friedman immediately set the statistical base for the creation of many algorithms, providing a general approach to boosting as optimization in the functional space.

In fact, there was a transition from engineering and algorithmic research in the construction of algorithms (so characteristic of ML) to a full-fledged methodology of how to build and study such algorithms. From the point of view of mathematical stuffing, at first glance, not much has changed: we all add (boost) weak algorithms, building up our ensemble with gradual improvements in those parts of the data where the previous models have not been finalized. But when constructing the next simple model, it is built not just on reweighted observations, but in such a way as to better approximate the overall gradient of the objective function. On a conceptual level, this gave great scope for imagination and expansion.

Gradient boosting did not take its place in the "gentleman's set"right away - it took more than 10 years from the moment it appeared (see Figure 3.8). Firstly, the base GBM has many extensions for various statistical tasks: GLMboost and GAMboost as a strengthening of existing GAM models, CoxBoost for survival curves, RankBoost and LambdaMART for ranking. Secondly, there are many implementations of the same GBM under different names and different platforms: Stochastic GBM, GBDT (Gradient Boosted Decision Trees), GBRT (Gradient Boosted Regression Trees), MART (Multiple Additive Regression Trees), GBM as Generalized Boosting Machines and others. In addition, the machine learner communities were quite fragmented and engaged in everything, because of this, tracking the success of boosting is quite difficult.

At the same time, boosting began to be actively used in ranking tasks for search engine results. This problem was written out in terms of the loss function, which fines for errors in the issuing order, so it became convenient to simply insert it into GBM. AltaVista was the first to introduce boosting in the ranking, and soon Yahoo, Yandex, Bing and others followed. Moreover, speaking about the introduction, it was said that boosting for years to come became the main algorithm inside the working engines, and not another interchangeable research work that lives in the framework of a couple of scientific articles.

The main role in popularizing boosting was played by ML competitions, especially kaggle. Researchers have long lacked a common platform where there would be enough participants and tasks for people to compete in their open struggle for state of the art with their algorithms and approaches. The gloomy German geniuses who had grown another miracle algorithm in their garage could no longer be attributed to closed data, and real breakthrough libraries, on the contrary, received an excellent platform for development. This is exactly what happened with the boost, which took root on kaggle almost immediately (you should look for GBM in an interview with the winners since 2011), and xgboost as a library quickly gained popularity shortly after its appearance. At the same time, xgboost is not some new unique algorithm, but simply an extremely effective implementation of the classic GBM with some additional heuristics.

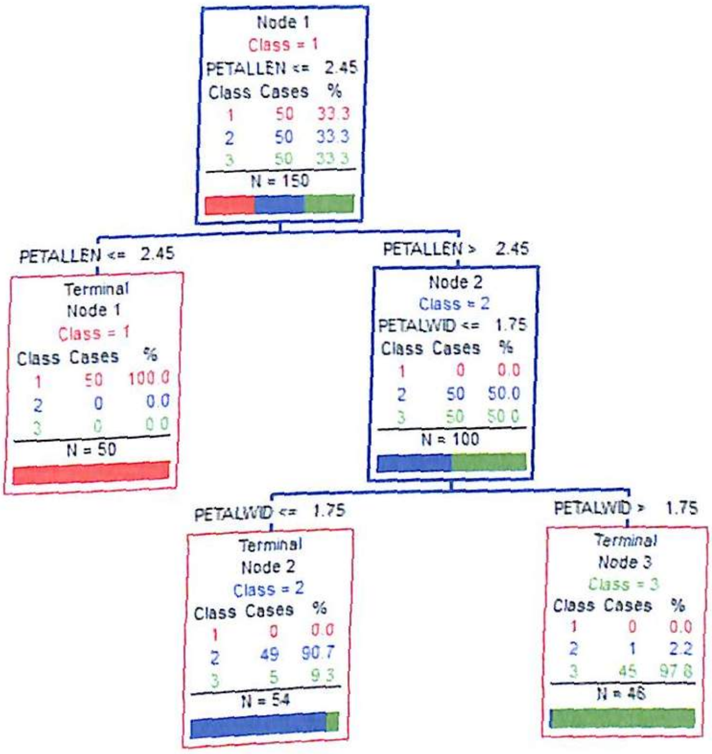

Figure. 3.9: Decision Tree

And here we are, in 2020, using an algorithm that went a very typical way for ML from a mathematical problem and algorithmic crafts through the emergence of normal algorithms and a normal methodology to successful practical applications and mass use years after its appearance.

Decision Tree

The decision tree is a tree in the leaves of which there are values of the objective function, and in the other nodes there are transition conditions (for example, "FLOOR is MEN'S") determining which edge to go on. If for this observation the condition is true, then the transition along the left edge is carried out, if the lie is on the right. The image above(Figure 3.9) shows the iris classification tree. The classification is divided into three classes (marked in the image – red, blue, and green), and it passes by the parameters: sepal length thickness (SepalLen, SepalWid) and petal length thickness (PetalLen, PetalWid). As you can see, each node has its belonging to the class (depending on which elements got into this node more), the number of observations that got there is N, as well as the number of each class. Also, not in leaf vertices there is a transition condition - to one of the child ones. Accordingly, according to these conditions, the sample is divided. As a result, this tree almost perfectly (6 out of 150 incorrectly) classified the initial data (namely, the initial data - those on which it was trained).

Regression

If the classification contains the resulting classes in the sheets, the regression is worth some value of the objective function.

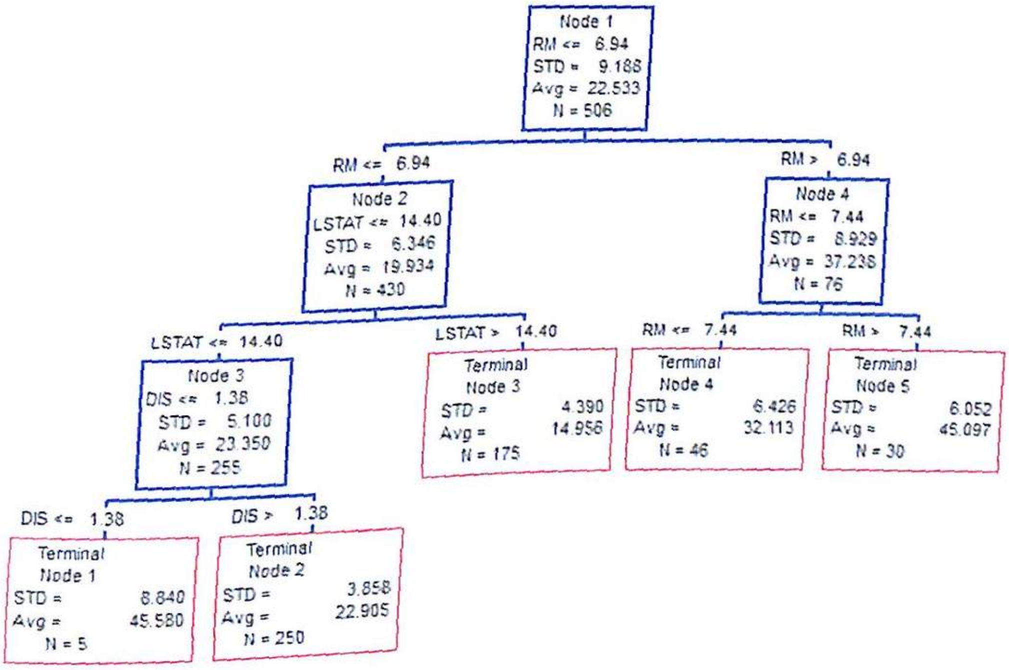

Figure. 3.10: Decision Tree

In the above image(Figure 3.10), a regression tree, to determine the price of land in the city of Boston in 1978, depending on the parameters RM - the number of rooms, LSTAT - percentage of the poor and several other parameters (for more details, see [4]). Accordingly, here in each node we see the mean value (Avg) and standard deviation (STD) of the values of the objective function of the observations that fell at this vertex. The total number of observations fell into node N. The result of the regression will be the average value (Avg) at which the observation will fall. Thus, the initial classification tree can also work for regression. However, with this approach, larger tree sizes are usually required than with classification in order to achieve good regression results.

Main methods

Listed below are a few basic methods that use decision trees. Their short description, pros and cons.

CART

CART (English Classification and regression trees - Classification and regression trees) was the first of the methods invented in 1983 by four famous scientists in the field of data analysis: Leo Breiman, Jerome Friedman, Richard Olshen and Stone. The essence of this algorithm is the usual construction of a decision tree, no more and no less. At the first iteration, we build all possible (in the discrete sense) hyperplanes that would split our space into two. For each such partition of space, the number of observations in each of the subspaces of different classes is considered. As a result, we choose a partition that maximally distinguishes one of the classes in one of the subspaces of observation. Accordingly, this partition will be our root of the decision tree, and the sheets at this iteration will be two partitions. At the next iterations, we take one worst (in terms of the ratio of the number of observations of different classes) sheet and carry out the same operation to split it. As a result, this sheet becomes a node with some kind of partition, and two sheets. We continue to do this until we reach the limit on the number of nodes, or from one iteration to another the general error (the number of incorrectly classified observations by the whole tree) ceases to improve. However, the resulting tree will be "retrained" (will be tailored to the training set) and, accordingly, will not give normal results on other data. In order to avoid "retraining", test samples (or cross-validation) are used and, accordingly, a reverse analysis is performed (the so-called pruning), when the tree is reduced depending on the result on the test sample. A relatively simple algorithm, which results in a single decision tree. Due to this, it is convenient for initial data analysis, for example, to check for the existence of relationships between variables and another. + Fast model building + Easy to interpret (due to the simplicity of the model, you can easily display the tree and trace all the nodes of the tree) - Often converges on a local solution (for example, in the first step a hyperplane was chosen that maximally divides the space at this step, but at the same time it will not lead to an optimal solution)

Random Forest Classifier

The concept of random forest was first introduced into scientific use in [14], [15], see also [16]. In these articles, many root forests with marked peaks were considered, on which a uniform probability distribution was specified. Later, a monograph appeared [17], in which random forests with distributions other than

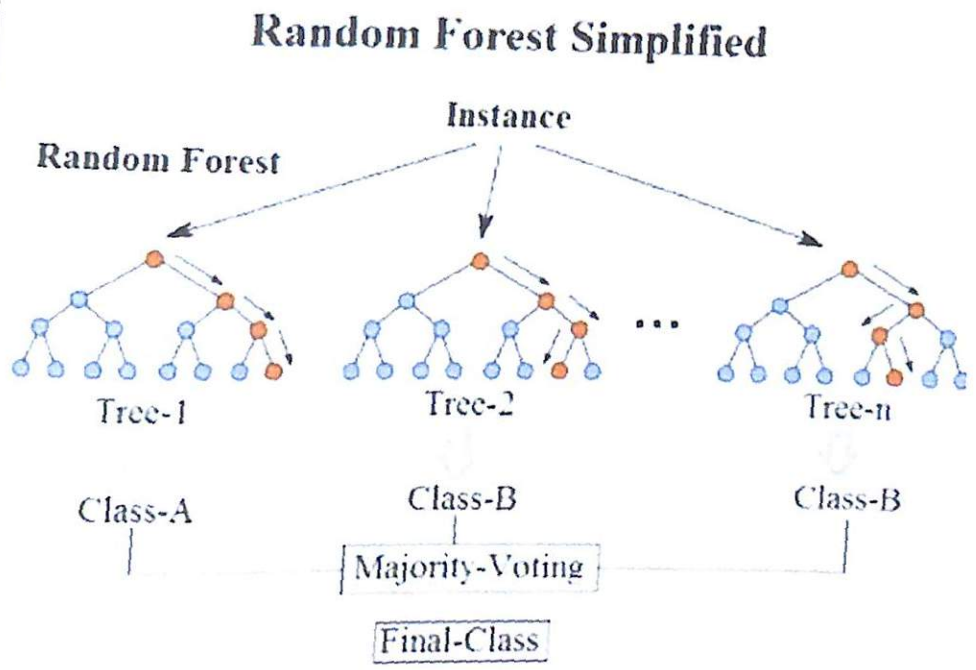

Figure. 3.11: Representation of Random Forest

uniform were studied. Thus, from the point of view of probability theory, random forests are a special case of the well-known concept of a random element (see [18]). However, in 2001, in article [19], a new method of classification and regression was proposed, also called a random forest. In this sense, the term random forest is widely used in disciplines such as machine learning, pattern recognition, a discipline known as "Data Mining" 1 and, to a lesser extent, in applied statistics. This review is dedicated to this method.

The method is based on the construction of a large number (ensemble) of decision trees (this number is a parameter of the method), each of which is constructed from a sample obtained from the initial training sample using a bootstrap (see Figure 3.11). In contrast to the classical algorithms for constructing decision trees [[20], [21]], in the random forest method, when constructing each tree at the stages of vertex splitting, only a fixed number of randomly selected features of the training sample are used (the second parameter of the method) and a complete tree is constructed (without truncation), e. Each leaf of the tree contains observations of only one class. Classification is carried out by voting of classifiers determined by individual trees, and regression is estimated by averaging the regression estimates of all trees. It is known (see, for example, [[22]]) that the accuracy (probability of correct classification) of classifier ensembles substantially depends on the diversity of classifiers that make up the ensemble or, in other words, on how correlated their solutions are.

Namely, the more diverse the classifiers of the ensemble (the less correlated their decisions), the higher the probability of a correct classification. In random forests, the solutions of their constituent trees are weakly correlated. Due to the double "injection of randomness" into the algorithm for constructing a random forest — at the bootstrap stage and at the stage of random selection of characters used in splitting tree tops.

The method quickly gained recognition both in the statistical community and among researchers using image recognition methods in their work and is currently one of the most popular methods of classification and nonparametric regression. The reason for this was not only the high accuracy of the classification (and, according to the author, not so much) provided by the method, but also its other advantages.

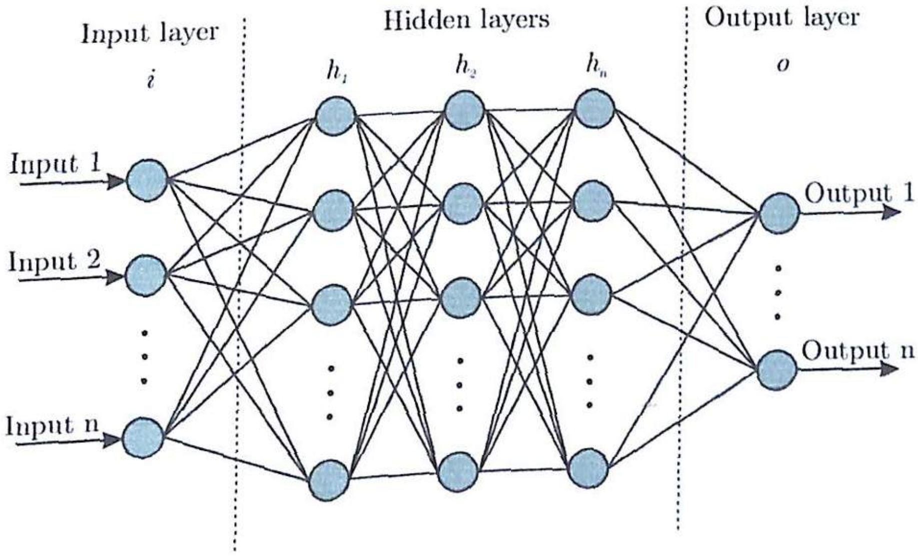

Figure. 3.12: Simple representation of Neural Network

Neural Network

Learning with a teacher assumes that for each input vector there is a target vector representing the desired output. Together they are called a training pair (see Figure 3.12). Typically, the network is trained on a number of such training pairs. An output vector is presented, the network output is calculated and compared with the corresponding target vector. Further, the weights are changed in accordance with an algorithm that seeks to minimize error. The vectors of the training set are presented sequentially, errors are calculated, and the weights are adjusted for each vector until the error over the entire training array reaches an acceptable level.

Weight setting networks with fixed connections - the weights of the neural network are selected immediately, based on the conditions of the problem; networks with dynamic connections - for them, in the learning process, the synaptic scales are set up. Input type analog - input information is presented in the form of real numbers; binary - all input information in such networks is represented in the form of zeros and ones. Applicable neural network model

Direct distribution networks - all connections are directed strictly from input neurons to output ones. Such networks include, for example: the simplest perceptron (developed by Rosenblatt) and a multilayer perceptron.

Recursive neural networks - the signal from the output neurons or neurons of the hidden layer is partially transmitted back to the inputs of the neurons of the input layer.

Radially basic functions - a type of neural network having a hidden layer of radial elements and an output layer of linear elements. Networks of this type are quite compact and quickly learn. Proposed by Broomhead and Lowe (1988) and Moody and Darkin (1989). The radially basic network has the following features: one hidden layer, only the hidden layer neurons have a nonlinear activation function and the synaptic weights of the input and hidden layers are equal to one.

Exactly:

- the method guarantees protection against overfitting 2 even when the number of signs significantly exceeds the number of observations. This property distinguishes the "random forest" method from many other classification methods and is extremely valuable for solving many applied problems;
- to build a random forest from the training set, only two parameters are required that require minimal tuning (tuning);
- The Out-Of-Bag (OOB) method proposed by Breiman [23] provides a natural estimate of the probability of an erroneous classification of random forests based on observations not included in the training bootstrap samples used to construct trees (these observations are called OOB samples);
- random forests can be used not only for classification and regression tasks, but also for the tasks of identifying the most informative features,

clustering, highlighting anomalous observations, and determining prototype classes;

- the training sample for constructing a random forest may contain features measured on different scales: numerical, ordinal, and nominal, which is unacceptable for many other classifiers;

1. There is no generally accepted translation of this term into Russian. Sometimes the literal translation "data mining" used does not hold water.

2. Re-fitting means a situation in which the classifier classifies the observations of the training sample well, but is unsuitable for classifying observations that are not included in it.

- the method allows easy parallelization (that is, a software implementation suitable for parallel computing), which is very important for large volumes of the training sample.

It seems to the author that the development of the random forest method took place in the following areas:

- study of the properties of the method itself, that is, analytical and experimental work on assessing accuracy, compared with other ensemble classification methods, etc.

- development of the capabilities of the method focused on solving problems that are not directly related to classification and regression problems (see above);

- development of other related methods based on the method, such as random survival forests, quantile regression forests, logical random forests, probabilistic random forests and stream random forests;

- the use of a random forest construction scheme to build ensembles of classifiers that are not trees – ensembles of naive Bayes classifiers and multinomial logistic models;

- development of algorithms and software that implement the method.

In accordance with this classification of studies, this review is built.

The first section is a brief excursion into the history of the method. The elements of the method are considered - decision trees, ensembles of classifiers, bug-ging (aggregated bootstrap) and the method of random subspaces. In this section, the author also considered it appropriate to briefly describe the basic concepts used in constructing decision trees — impurity, split, and pruning of decision trees.

Naive Bayes Classification

A naive Bayesian classifier is a family of classification algorithms that take one assumption: Each parameter of the classified data is considered independently of other class parameters.Why is the method called naive? The assumption that all data set parameters are independent is a rather naive assumption. This usually doesn't happen.

What does the word "independently" mean? 2 parameters are called independent when the value of one parameter does not affect the second. For instance: Say you have a patient data set: pulse, cholesterol, weight, height, and zip code. All parameters will be independent if the values of all parameters do not affect each other. For this data set, it is reasonable to assume that the patient's height and zip code are independent, as the person's height and zip code are not related. But let's look further, are all the parameters independent? Unfortunately, the answer is no. There are 3 relationships that are dependent:

- if height increased, weight probably increased;
- if cholesterol increased, weight probably increased;
- if cholesterol has increased, the pulse is likely to increase.

Typically, data set parameters are not completely independent. Why is the method called naive? The assumption that all data set parameters are independent is a rather naive assumption. This usually doesn't happen.

Who is Bayes? Thomas Bayes was an English statistician and mathematician named after Bayes' Theorem. In fact, the theorem allows us to predict the class based on a set of parameters using probability. A simplified equation for classification looks like this (see Figure 3.13):

$$P(Class\ A|Feature\ 1, Feature\ 2) = \frac{P(Feature\ 1|Class\ A) \cdot P(Feature\ 2|Class\ A) \cdot P(Class\ A)}{P(Feature\ 1) \cdot P(Feature\ 2)}$$

Figure. 3.13: Naive Bayes Algorithm

Let's take a look at it in more detail. What does this equation mean? The equation finds the probability of class A, based on parameters 1 and 2. In other words, if you see parameters 1 and 2, then this is probably the data of class A.

The equation reads as follows: The probability of [identifying] Class A based on parameters 1 and 2 is a fraction. The fraction numerator is the probability of parameter 1 belonging to class A, multiplied by the probability of parameter 2, belonging to class A, multiplied by the probability of class A. The denominator is the probability of parameter 1 times the probability of parameter 2.

3.3 Implementation of models

The process by which companies create value for customers and build strong customer relationships in order to capture value from customers in return. Marketing campaigns are characterized by focusing on the customer needs and their overall satisfaction. Nevertheless, there are different variables that determine whether a marketing campaign will be successful or not. There are certain variables that we need to take into consideration when making a marketing campaign.

1) Segment of the Population: To which segment of the population is the marketing campaign going to address and why? This aspect of the marketing campaign is extremely important since it will tell to which part of the population should most likely receive the message of the marketing campaign.

2) Distribution channel to reach the customer's place: Implementing the most effective strategy in order to get the most out of this marketing campaign. What segment of the population should we address? Which instrument should we use to get our message out? (Ex: Telephones, Radio, TV, Social Media Etc.)

3) Price: What is the best price to offer to potential clients? (In the case of the bank's marketing campaign this is not necessary since the main interest for the bank is for potential clients to open deposit accounts in order to make the operative activities of the bank to keep on running.)

4) Promotional Strategy: This is the way the strategy is going to be implemented and how are potential clients going to be address. This should be the last part

of the marketing campaign analysis since there has to be an indepth analysis of previous campaigns (If possible) in order to learn from previous mistakes and to determine how to make the marketing campaign much more effective.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import squarify
import matplotlib.pyplot as plt

from plotly import tools
import chart_studio.plotly as py
import plotly.figure_factory as ff
import plotly.graph_objs as go
from plotly.offline import download_plotlyjs, init_notebook_mode, plot, iplot
```

Figure. 3.14: List of used libraries for Data representation

To work with giant amount data and for see the nice representation of them, I have imported different kind of libraries also basics to operate with them (see Figure 3.14).

The matplotlib library is a two-dimensional graphics library for the python programming language with which you can create high-quality drawings of various formats. Matplotlib is a module package for python. Matplotlib consists of many modules. Modules are filled with various classes and functions that are hierarchically related.

Seaborn is essentially a higher-level API based on the matplotlib library. Seaborn contains more appropriate default charting settings. Also in the library there are quite complex types of visualization that in matplotlib would require a lot of code.

Plotly is positioned as an online platform where you can create and publish your own charts. However, this library can be used simply in Jupyter Notebook'e. In addition, the library has offline-mode, which allows you to use it without registering and publishing data and graphs to the plotly server[23].

In general, I really liked the library: there is detailed documentation with examples, various types of graphs are supported (scatter plots, box plots, 3D graphs, bar charts, heatmaps, dendrograms, etc.) and the graphs are pretty nice.

A Term deposit is a deposit that a bank or a financial institurion offers with a fixed rate (often better than just opening deposit account) in which your money will be returned back at a specific maturity time. For more information with regards to Term Deposits please click on this link from Investopedia [24].

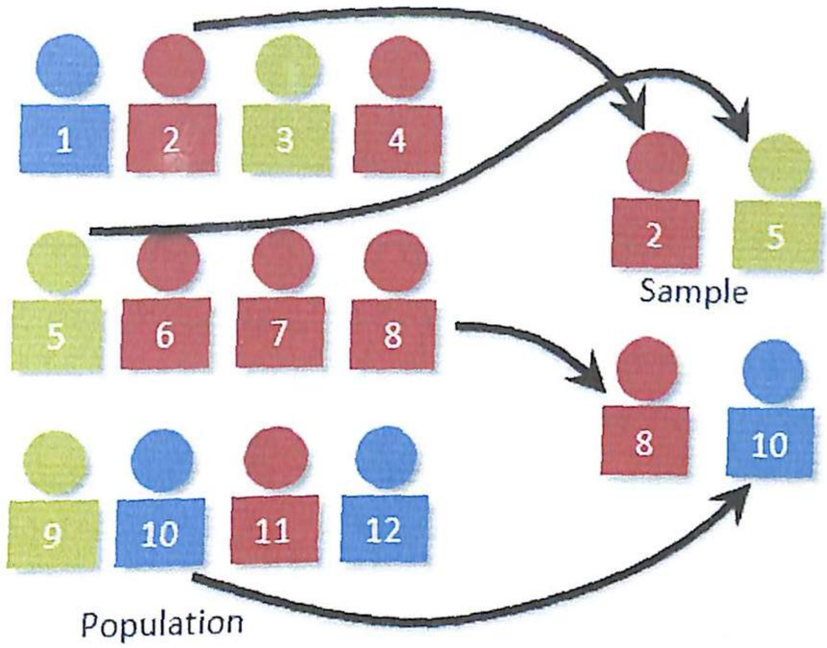

Figure. 3.15: Stratified Sampling implementation

Stratified Sampling - Is a significant idea that is regularly missed when building up a model either for relapse or *grouping (see* Figure 3.15). Keep in mind, that so as to abstain from overfitting of our information we should actualize a cross approval in any case, we should ensure that in any event the highlights that have the best impact on our mark (regardless of whether a potential customer will open a term store or not) is similarly circulated. I'm not catching my meaning by this?

For example, having an individual advance is a significant component that decides if a potential customer will open a term store or not. To affirm it has an overwhelming load on the last yield you can check the relationship grid above and you can see it has a - 11% connection with opening a store. What steps we should take before executing stratified examining in our train and test information?

1) We have to perceive how our information is disseminated.

2) After noticiing that the segment of credit contains 87% of "no"(Does not have individual advances) and 13% of "yes"(Have individual advances.)

3) We need to ensure that our preparation and test set contains a similar proportion of 87% "no"and 13% "yes"."Stratified Sampling: Is a significant idea that is frequently missed when building up a model either for relapse or grouping. Keep in mind, that so as to abstain from overfitting of our information we should execute a cross approval nonetheless, we should ensure that at any rate the highlights that have the best impact on our name (regardless of whether a potential

customer will open a term store or not) is similarly appropriated. I don't get my meaning by this?

Now we split the data into training and test sets and implement a stratified shuffle split. Data we will separate by usual proportion like 0.8 to train and 0.2 to test. Also need to separate the labels and the features.

CategoricalEncoder class that we use, to convert categories that presented as text would be tranformed to numbers to operate on them. Encode straight out highlights as a numeric exhibit. The contribution to this transformer ought to be a lattice of numbers or strings, signifying the qualities taken on by absolute (discrete) highlights. The highlights can be encoded utilizing a one-hot otherwise known as one-of-K conspire ("encoding='onehot"', the default) or changed over to ordinal whole numbers ("encoding='ordinal"'). This encoding is required for taking care of clear cut information to numerous scikit-learn estimators, outstandingly direct models and SVMs with the standard pieces.

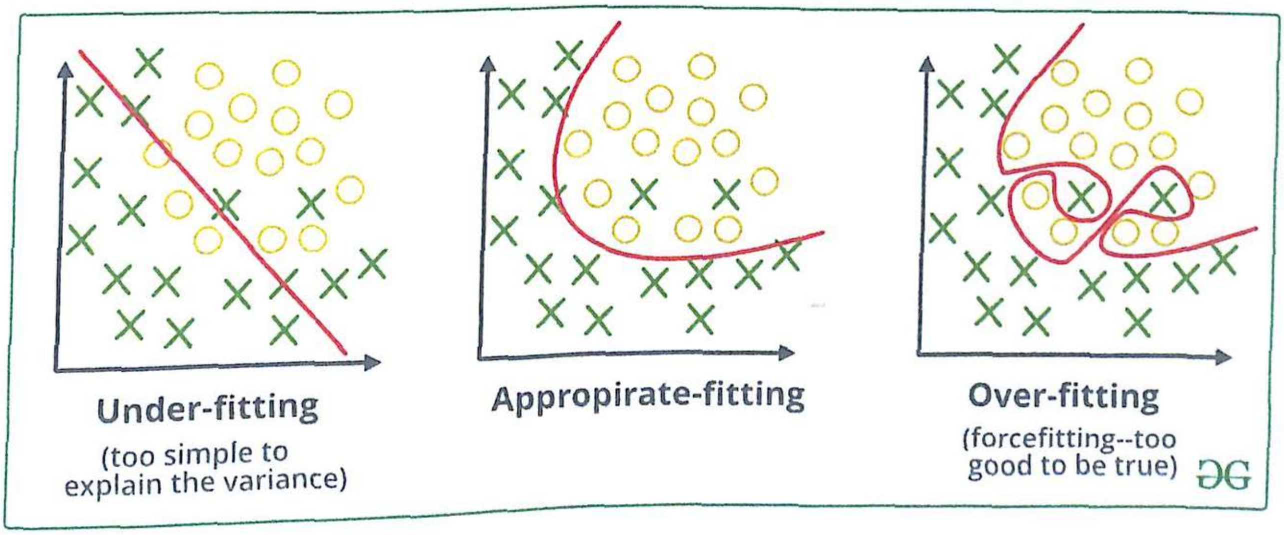

Figure. 3.16: Example and difference of overfitting

Parameters is following: - encoding : str, 'onehot', 'onehot-thick' or 'ordinal' The sort of encoding to utilize (default is 'onehot'): - 'onehot': encode the highlights utilizing a one-hot otherwise known as one-of-K plot (or on the other hand additionally called 'sham' encoding). This makes a double section for every class and returns an inadequate grid. - 'onehot-thick': equivalent to 'onehot' yet restores a thick exhibit rather than a meager grid. - 'ordinal': encode the highlights as ordinal numbers.

Avoiding Overfitting:

Brief Description of Overfitting?

Running Time of classifiers		
Name	Train Score	Training Time
Decision Tree	1.000000	0.091520
Random Forest	0.997424	0.234199
Nearest Neighbors	0.863255	0.241870
Gradient Boosting Classifier	0.861463	1.655768
Neural Net	0.853735	11.977817
Linear SVM	0.852391	3.959715
Logistic Regression	0.830776	0.055697
Naive Bayes	0.721693	0.033629

Table. 3.4: Table of countries

This is a blunder in the demonstrating calculation that contemplates arbitrary clamor in the fitting procedure as opposed to the example itself. You can see this happens when the model gets an awesome score in the preparation set however when we utilize the test set (Unknown information for the model) we get a dreadful score. This is probably going to happen in view of overfitting of the information (thinking about irregular commotion in our example). What we need our model to do is to take the general example of the information so as to effectively arrange whether a potential customer will subscribe to a term store or not. In the models above, all things considered, the Decision Tree Classifier and Random Forest classifiers are overfitting since the two of them give us almost immaculate scores (100% and 99%) precision scores.

And below is time spend and train score for each classification (see Table 3.4)

How might we abstain from Overfitting?

The best choice to abstain from overfitting is to utilize cross approval. Taking the preparation test and parting it. For example, in the event that we split it by 3, 2/3 of the information or 66% will be utilized for preparing and 1/3 33% will be utilized or testing and we will do the testing procedure multiple times. This calculation will repeat through all the preparation and test sets and the principle reason for this is to get the general example of the information. After applying crossvalidation the scores have changed(see Table 3.5).

Bits of knowledge of a Confusion Matrix:

The principle reason for a disarray framework is to perceive how our model is performing with regards to grouping potential customers that are probably going

Classifiers crossval Mean Scores	
Name	Score
Decision Tree	0.786313
Random Forest	0.843655
Nearest Neighbors	0.804458
Gradient Boosting Classifier	0.845224
Neural Net	0.847689
Linear SVM	0.840186
Logistic Regression	0.828425
Naive Bayes	0.847689

Table. 3.5: Table of countries

		Prediction	
		0	1
Actual	0	TN	FP
	1	FN	TP

Figure. 3.17: Confusion Matrix

to subscribe to a term store. We will find in the disarray grid four terms the True Positives, False Positives, True Negatives and False Negatives(see Figure 3.17).

Positive/Negative: Type of Class (mark) ["No "Yes"] True/False: Correctly or Incorrectly grouped by the model.

Genuine Negatives (Top-Left Square): This is the quantity of effectively groupings of the "No"class or potential customers that are not willing to subscribe a term store.

Bogus Negatives (Top-Right Square): This is the quantity of erroneously groupings of the "No"class or potential customers that are not willing to subscribe a term deposit.

Bogus Positives (Bottom-Left Square): This is the quantity of erroneously groupings of the "Yes"class or potential customers that are willing to subscribe a term store.

Genuine Positives (Bottom-Right Square): This is the quantity of effectively groupings of the "Yes"class or potential customers that are willing to subscribe a term store.

Precision and Recall:

Review: Is the all out number of "Yes"in the mark section of the dataset. So what number of "Yes"marks does our model recognize.

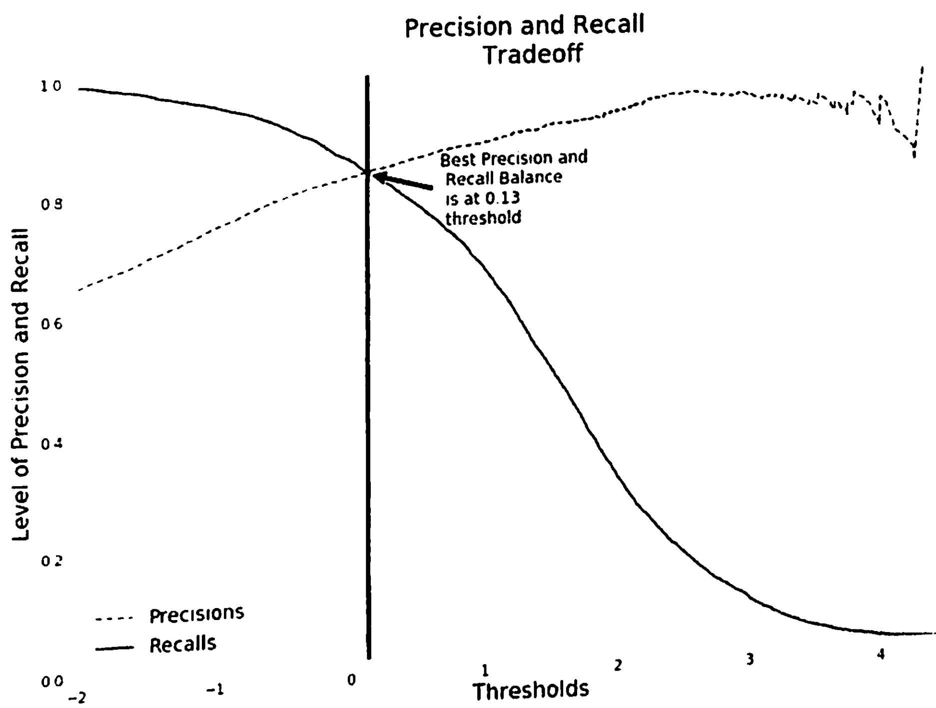

Figure. 3.18: Precision and Recall trade offs

Precision: Means how sure is the expectation of our model that the real mark is a "Yes".

Review Precision Trade off:

As the accuracy gets higher the review gets lower and the other way around. For example, in the event that we increment the exactness from 30% to 60% the model is picking the forecasts that the model accepts is 60% certain. On the off chance that there is a case where the model accepts that is 58% prone to be a potential customer that will subscribe to a term store then the model will arrange it as a "No."However, that example was really a "Yes"(potential customer did subscribe to a term store.) That is the reason the higher the exactness the more probable the model is to miss occasions that are really a "Yes"(see Figure 3.18)!

Precision Score: 0.8244135732179458

Recall Score: 0.8553875236294896

ROC Curve (Receiver Operating Characteristic):

The ROC bend discloses to us how well our classifier is grouping between term store subscriptions (True Positives) and non-term store subscriptions. The X-pivot is spoken to by False positive rates (Specificity) and the Y-hub is spoken to by the True Positive Rate (Sensitivity.) As the line moves the edge of the order changes giving us various qualities. The closer is the line to our upper left corner

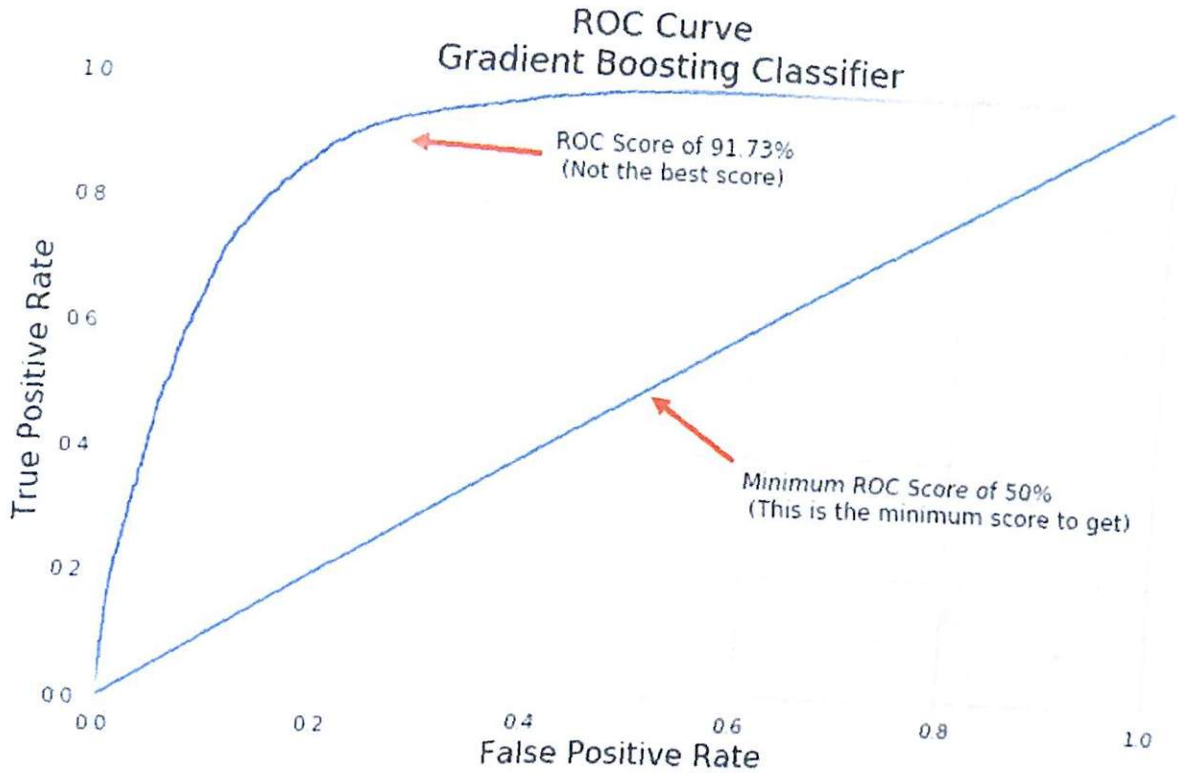

Figure. 3.19: Receiver Operating Characteristic of Gradient Boosting classifier

the better is our model isolating the two classes (see Figure 3.19).

To compare with others (see Figure 3.20):

Gradient Boost Classifier Score: 0.9173128596743366

Neural Classifier Score: 0.9167698643666292

Naives Bayes Classifier: 0.803363959942255

Which Features Influence the Result of a Term Deposit Subscription?

DecisionTreeClassifier:

The best three most significant highlights for our classifier are **Duration (to what extent it took the discussion between the agent and the potential customer), contact (number of contacts to the potential customer inside a similar showcasing effort), month (the period of the year).

GradientBoosting Classifier Wins!

Inclination Boosting classifier is the best model to anticipate whether a potential customer will subscribe to a term store or not. 84% exactness! By consolidating every one of these methodologies and rearranging the market crowd the following effort should address, almost certainly, the following promoting effort of the bank will be more powerful than the present one. If you go in more detail on what attributes are worth paying attention to, this is the following:

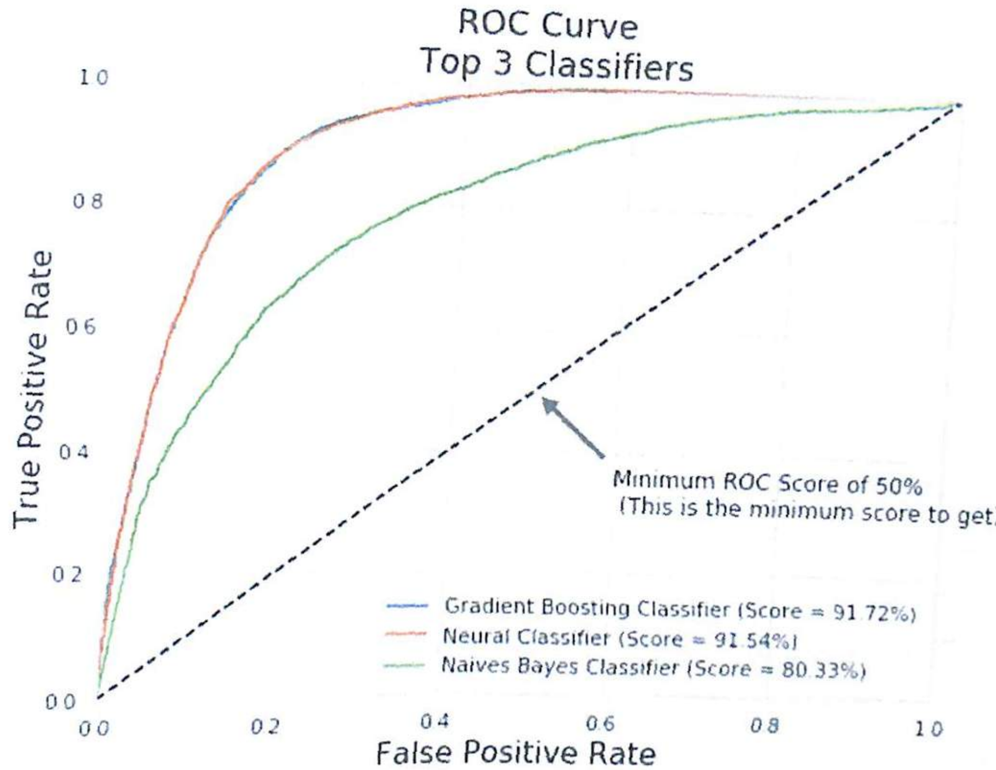

Figure. 3.20: ROC 3 top classifiers

1) Months of Marketing Activity: We saw that the period of most elevated level of showcasing movement was the long stretch of May. In any case, this was the month that potential customers would in general reject term stores offers (Lowest compelling rate: - 34.49%). For the following showcasing effort, it will be shrewd for the bank to center the promoting effort during the long periods of March, September, October and December. (December ought to be getting looked at in light of the fact that it was the month with the most reduced showcasing action, there may be a motivation behind why december is the least.)

2) Seasonality: Potential customers selected to suscribe term stores during the periods of fall and winter. The following showcasing effort should center its movement throghout these seasons.

3) Campaign Calls: An approach ought to be actualized that expresses that close to 3 calls ought to be applied to a similar potential customer so as to spare time and exertion in getting new potential customers. Keep in mind, the more we call a similar potential customer, the presumable the person will decay to open a term store.

4) Age Category: The following showcasing effort of the bank should target potential customers in their 20s or more youthful and 60s or more seasoned. The

most youthful classification had a 60% possibility of suscribing to a term store while the oldest class had a 76% possibility of suscribing to a term store. It will be extraordinary if for the following effort the bank tended to these two classes and in this way, improve the probability of more term stores suscriptions.

5) Occupation: of course, potential customers that were understudies or resigned were the destined to suscribe to a term store. Resigned people, will in general have more term stores so as to increase some money through premium installments. Keep in mind, term stores are momentary advances in which the person (right now resigned individual) makes a deal to avoid pulling back the money from the bank until a specific date concurred between the individual and the monetary establishment. After that time the individual recovers its capital and its advantage made on the credit. Resigned people tend to not go through bigly its money so they are morelikely to give their money something to do by loaning it to the budgetary organization. Understudies were the other gathering that used to suscribe term stores.

6) House Loans and Balances: Potential customers in the low adjust and no parity class were bound to have a house credit than individuals in the average and high parity classification. I'm not catching it's meaning to have a house credit? This implies the potential customer has budgetary trade offs to repay its home advance and along these lines, there is no money for the person to suscribe to a term store account. Be that as it may, we see that potential customers in the normal and hih balances are less inclined to have a house credit and subsequently, bound to open a term store. Ultimately, the following advertising effort should concentrate on people of normal and high adjusts so as to improve the probability of suscribing to a term store.

7) Develop a Questionaire during the Calls: Since span of the call is the component that most decidedly relates with whether a potential customer will open a term store or not, by giving a fascinating questionaire to potential customers during the calls the discussion length may increment. Obviously, this doesn't guarantee us that the potential customer will suscribe to a term store! In any case, we don't free anything by actualizing a procedure that will expand the degree of commitment of the potential customer prompting an expansion likelihood of suscribing to a term store, and in this way an expansion in viability for the following promoting effort the bank will excecute.

8) Target people with a higher span (over 375): Target the objective gathering that is better than expected in length, there is an exceptionally probability that this objective gathering would open a term store account. The probability that this gathering would open a term store account is at 78% which is truly high. This would permit that the achievement pace of the following promoting effort would be profoundly effective.

4. Conclusion

Using the NBA method, we achieved exactly what we wanted, namely an individual approach to each client.

The data used for the analysis is not bad, it would be better if there would be more in number than 48000 and in properties than 15. It would be interesting to observe the correlation of the influence of each attribute on the possible outcome of the data, so it will be easy to say for a large number data what approximate result we expect or in what range it will be.

Most of the selected classifiers did their job, but two of them showed great retraining of the Decision Tree Classifier and Random Forest classifiers (100% and 99%) precision scores, which meant maybe the others showed inaccurate results too. The solution was to add a cross validation method to remove overfitting and also show the final results more accurately. The confusion matrix method was also used - to see how our model works when it comes to classifying potential customers who can count on a term deposit. As the last step, the ROC Curve (Receiver Operating Characteristic) was used to identify the most accurate classifier.

In conclusion, we can say that all the selected classifiers in one sense or another helped for the final result. Nevertheless, the Gradient Boosting Model with an accuracy of 84% became the leader for our purposes and for this kind of data, it also proved to be very good in speed, but it was not the fastest like the Naive Bayes.

References

[1] (2015). Veripark's solution, url: https://www.veripark.com/products/veriloan. (accessed: 01.03.2020).

[2] (2018). RapidMiner's solution, url: https://rapidminer.com/solutions/next-best-action/. (accessed: 01.03.2020).

[3] (2016). CloudSense's solution, url: https://www.cloudsense.com/platform/genius-next-best-action. (accessed: 01.03.2020).

[4] (2017). Pega's solution, url: https://www.pega.com/technology/next-best-action. (accessed: 01.03.2020).

[5] (2015). NGData's solutions, url: https://www.ngdata.com/solutions/. (accessed: 01.03.2020).

[6] (2017). Jacada's solution, url: https://www.jacada.com/solutions/agent-next-best-action. (accessed: 01.03.2020).

[7] Ч. О, «Special-project in SBERBANK», *IT Community*, c. 1, Jan 30, 2018. url: https://habr.com/company/jugru/blog/347854/.

[8] M. A. N. Rory P. Bunker Wenjun Zhang. (1 Mar, 2017). Improving a Credit Scoring Model by Incorporating Bank Statement Derived Features, url: https://arxiv.org/abs/1611.00252v2. (accessed: 01.03.2020).

[9] V. S. K. F. I. S. D. Berestnev2 и M. Panov1. (23 Jan, 2020). Linking Bank Clients using Graph Neural Networks Powered by Rich Transactional Data, url: https://arxiv.org/pdf/2001.08427v1.pdf. (accessed: 01.05.2020).

[10] M. W. |. N. D.-T. |. N. S. | и L. Schmidt-Thieme. (13 Oct, 2016). Bank *Card Usage Prediction Exploiting Geolocation Information*, url: https://arxiv.org/pdf/1610.03996v1.pdf. (accessed: 01.05.2020).

[11] X. W. Seyed Mehdi Ayyoubzadeh. (26 Nov, 2019). Filter Bank Regularization of Convolutional Neural Networks, url: `https://arxiv.org/pdf/1907.11110v3.pdf`. (accessed: 01.05.2020).

[12] (2012). Vector model, url: `http://www.machinelearning.ru/wiki/index.php?title=%5C%D0%5C%92%5C%D0%5C%B5%5C%D0%5C%BA%5C%D1%5C%82%5C%D0%5C%BE%5C%D1%5C%80%5C%D0%5C%BD%5C%D0%5C%B0%5C%D1%5C%8F_%5C%D0%5C%BC%5C%D0%5C%BE%5C%D0%5C%B4%5C%D0%5C%B5%5C%D0%5C%BB%5C%D1%5C%8C`. accessed: Mar 16, 2020.

[13] (2013). Detecting Insults in Social Commentary Task, url: `https://www.kaggle.com/c/detecting-insults-in-social-commentary`. accessed: Mar 16, 2020.

[14] П. Ю. Л., *Limit theorems for the number of trees of a given volume in a random forest*, сер. Mathematical Collection 3. 1977, т. 103, с. 392—403.

[15] ——, *Asymptotic distribution of the maximum volume of a tree in a random forest*, сер. Mathematical Collection 3. 1977, т. 22, с. 523—533.

[16] ——, *Probability and math statistics*. Big Russian Encyclopedia, 1999, с. 604—605.

[17] P. Y. L., *Random Forests*. 2000, с. 122.

[18] K. B. Ф., *Random Mappings*, сер. Science. 1984, 208a.

[19] B. L., *Random forests*, сер. Machine Learning 1. 2001, т. 45, с. 5—32.

[20] B. L. F. R. O. R. S. C., *Classification and Regression Trees*. 1984, с. 342.

[21] Q. J. R., *Simplifying decision trees*, сер. International Journal of ManMachine Studies. 1987, т. 22, с. 221—234.

[22] K. L. I., *Combining Pattern Classifiers: Methods and Algorithms*. 2004, с. 349.

[23] P. community. (). Plotly documentation, url: `https://plotly.com/python/renderers/`. (accessed: 01.05.2020).

[24] (2020). Term Deposit definition, url: `https://www.investopedia.com/terms/t/termdeposit.asp`. accessed: March 16, 2020.

A. Appendix A

A.1 Data graph representation

Balance by Occupation: Management and Retirees are the ones who have the most elevated equalization in their records. (Figure A.1)

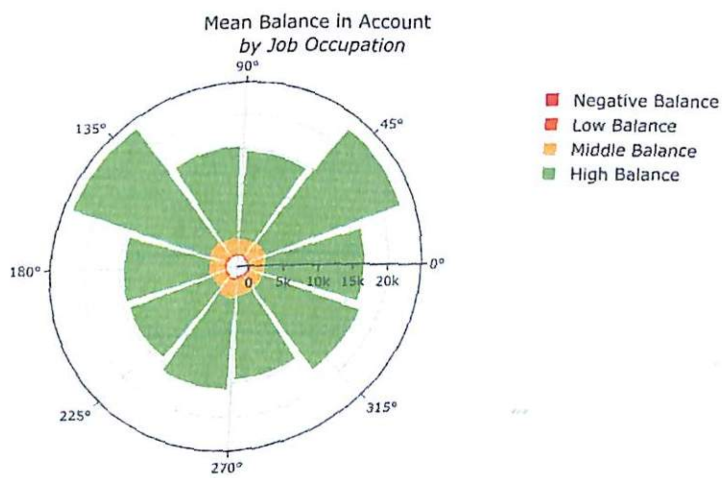

Figure. A.1: Balance by Occupation

Age Distribution (see Figure A.3)

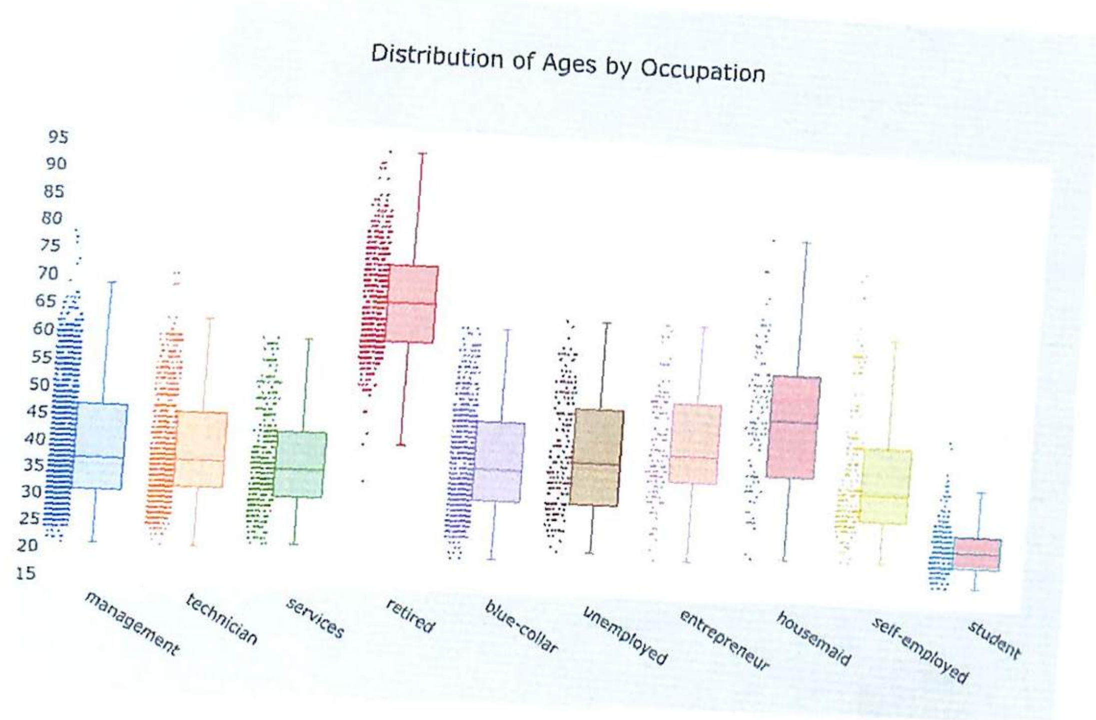

Figure. A.2: Age by Occupation

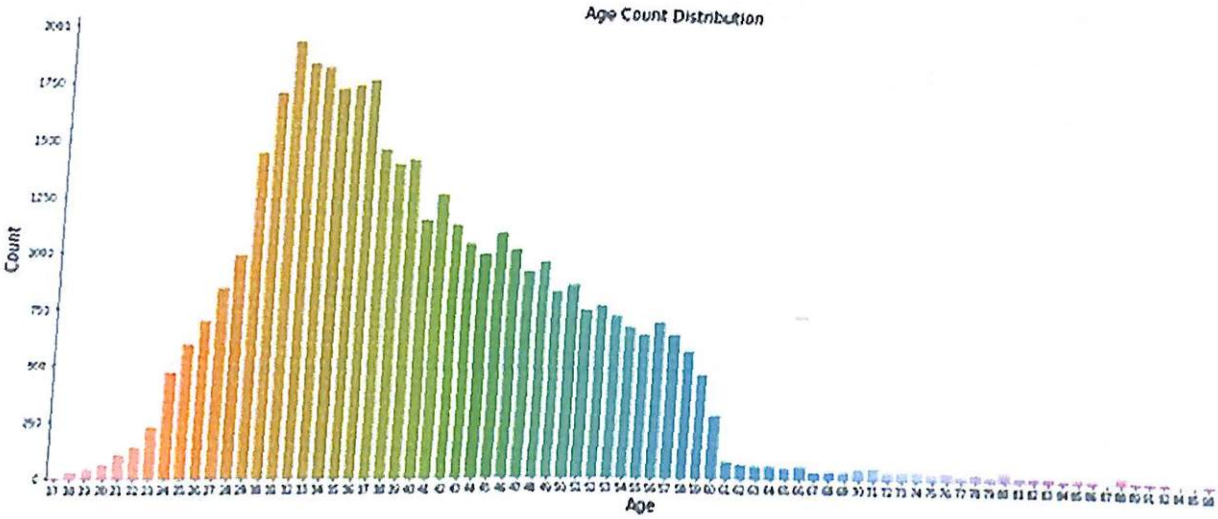

Figure. A.3: Age Distribution

Job Distribution (see Figure A.4)
Education Distribution (see Figure A.5)
Marital Distribution (see Figure A.6)

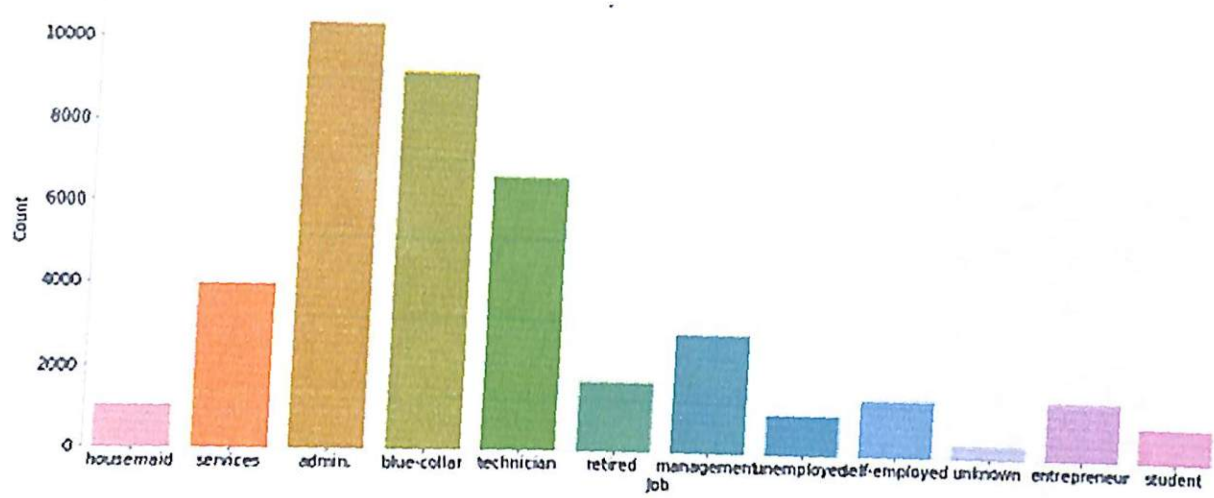

Figure. A.4: Job Distribution

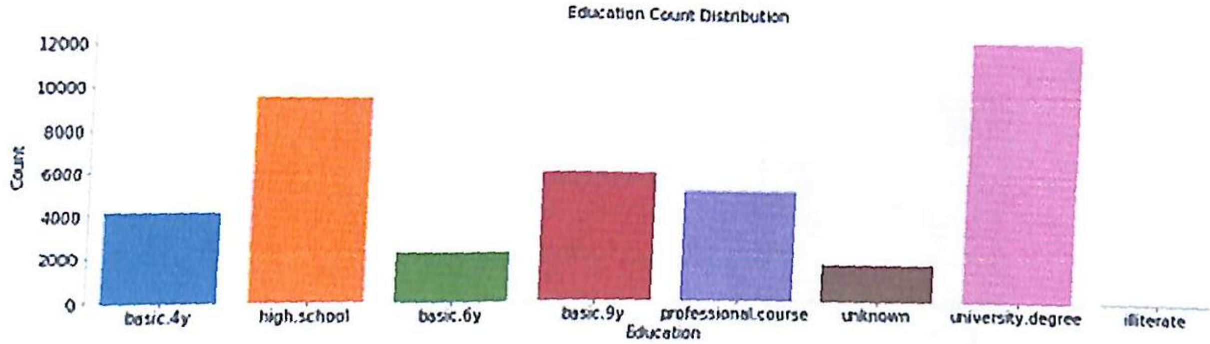

Figure. A.5: Education Distribution

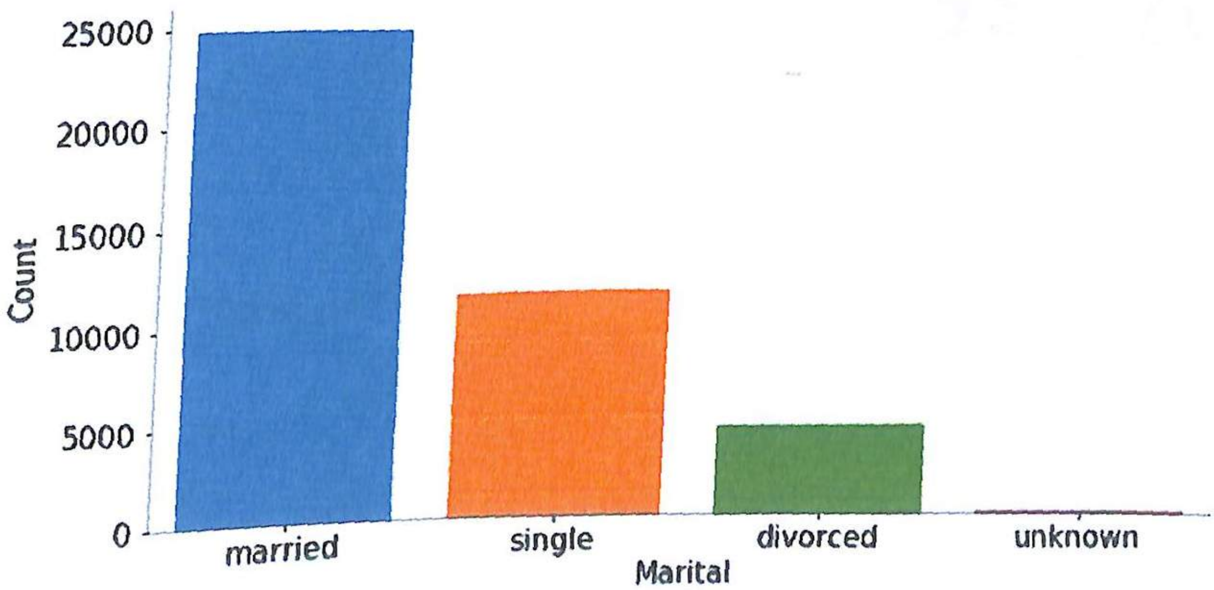

Figure. A.6: Marital Distribution

Contacts counts and Month counts and Day of week counts (see Figure A.7)
Default and Hosing and Loan (see Figure A.8)

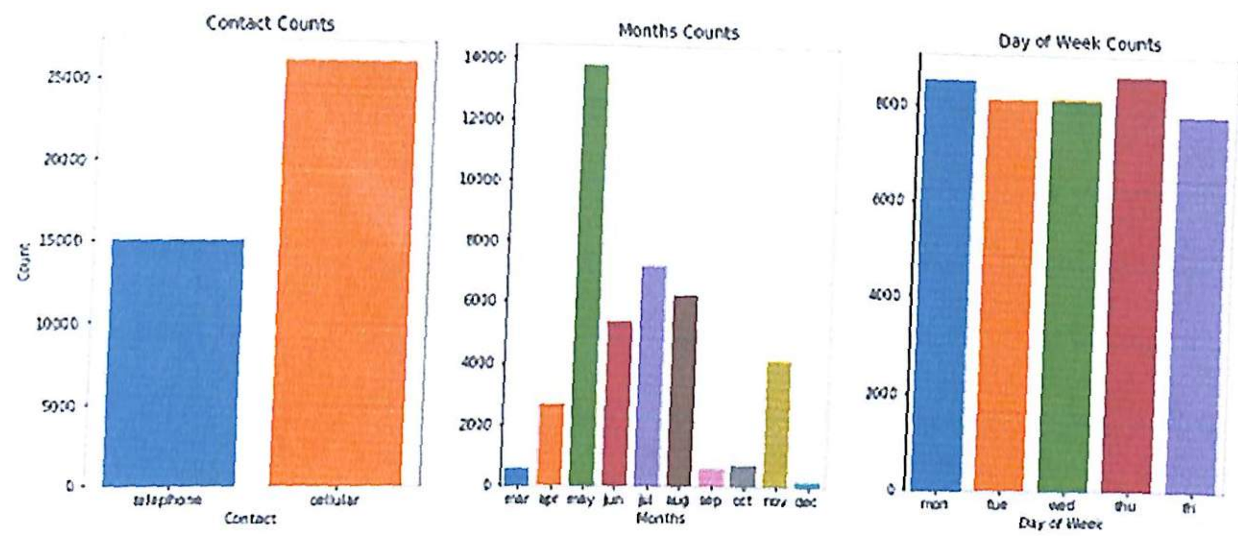

Figure. A.7: Contacts counts and Month counts and Day of week counts

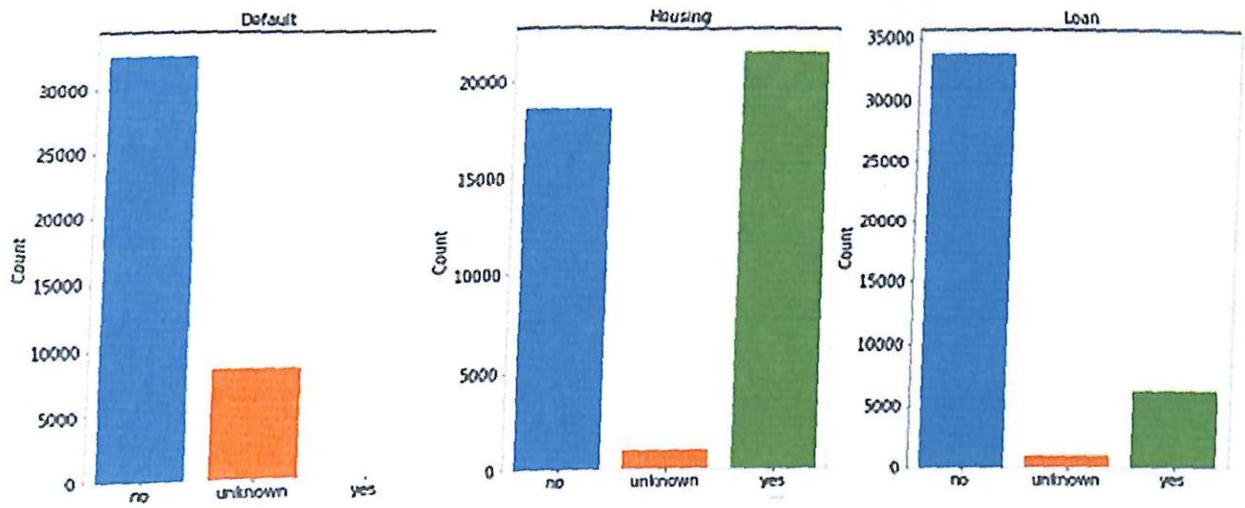

Figure. A.8: Default and Hosing and Loan

Count by marital status (see Figure A.9)

Balance by duration (see Figure A.10)

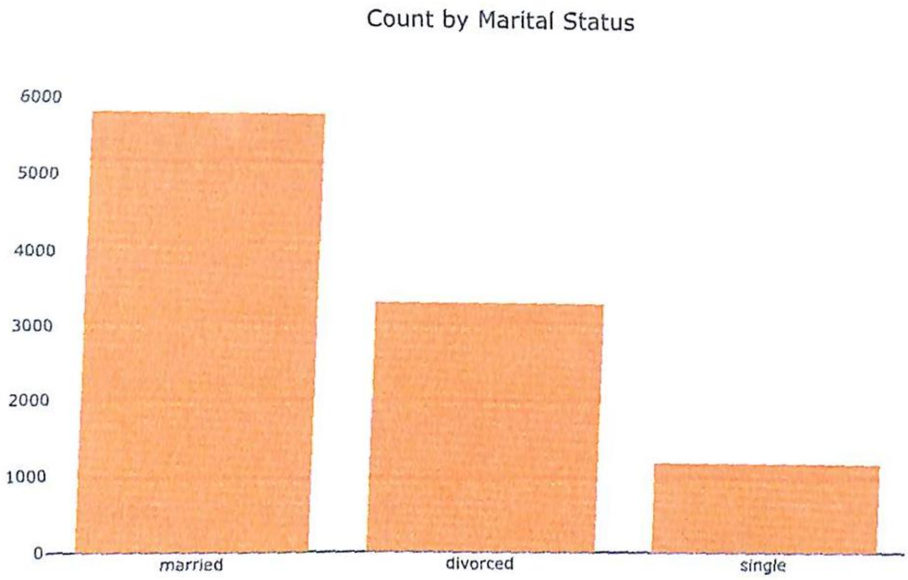

Figure. A.9: Count by marital status

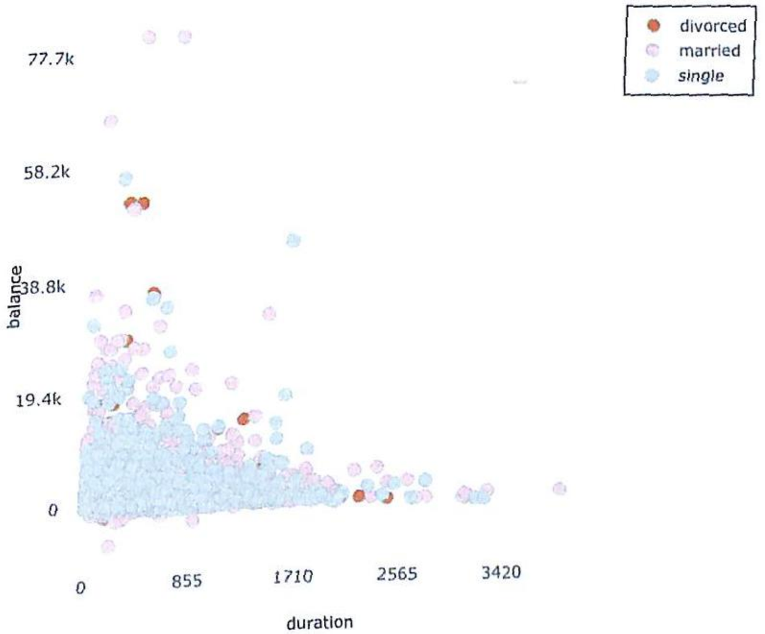

Figure. A.10: Balance and Duration

B. Appendix B

```python
# Time for Classification Models
import time


from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler, LabelEncoder

from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier
from sklearn import tree
from sklearn.neural_network import MLPClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.gaussian_process.kernels import RBF
from sklearn.ensemble import RandomForestClassifier
from sklearn.naive_bayes import GaussianNB


dict_classifiers = {
    "Logistic Regression": LogisticRegression(),
    "Nearest Neighbors": KNeighborsClassifier(),
    "Linear SVM": SVC(),
    "Gradient Boosting Classifier": GradientBoostingClassifier(),
    "Decision Tree": tree.DecisionTreeClassifier(),
    "Random Forest": RandomForestClassifier(n_estimators=18),
    "Neural Net": MLPClassifier(alpha=1),
    "Naive Bayes": GaussianNB()
}
```

Figure. B.1: Import and implement classifications with parameters