

Student Career Prediction Using Algorithms Of Machine Learning

Ankita Sinha
Department of CSE
Chandigarh University
Mohali, India
19BCS2354@cuchd.in

Garima
Department of CSE
Chandigarh University
Mohali, India
19BCS2879@cuchd.in

Ashish Singh
Department of CSE
Chandigarh University
Mohali, India
19BCS2915@cuchd.in

Abstract— In this day and age, choosing a career is hard work. Most students do not understand the concept of work from the early stages of their lives because they lack the development and knowledge of business. In addition, students are highly influenced when deciding which career path will lead to the best results. Students do not have enough information to make decisions on their own, which can cause problems in the future. To avoid this problem in the future, students should make good decisions when choosing the highest paying job for them. Choosing the wrong job that is not for them, they may end up in a job they don't like or they don't have much experience in that job. Since students are indecisive, they seek fortune tellers, hoping they can guide them towards the future. Do not rely on fortune tellers to make the best predictions about the future. Considering all this in this study, we will explore and explore the effectiveness of predicting career opportunities through research data. The model will suggest career options to students based on their abilities and performance in their careers. If students finish their studies with skill and performance, they can choose that field, otherwise they will have to leave that field and move on to another field. This article provides a career prediction using machine learning to help students choose the right career for their future.

Career forecasting is an important task in education. In this article, we introduce machine learning as a method for predicting student performance. We collect data from many sources, including education, interests, and personality, and use a combination of supervised and unsupervised machine learning algorithms to predict our best wishes to all students. If students finish their studies with skill and performance, they can choose that field, otherwise they will have to leave that field and move on to another field. This article provides a career prediction using machine learning to help students choose the right

career for their future.

Career forecasting is an important task in education. In this article, we introduce machine learning as a method for predicting student performance. We collect data from many sources, including education, interests, and personality, and use a combination of supervised and unsupervised machine learning algorithms to predict our best wishes to all students.

Keywords: Capabilities, Logistic regression, svm, classification, Machine learning, Neural network.

Introduction

Traditionally, student performance can be estimated using questionnaires. However, this method takes a lot of time. Computers now play an important role in many fields. Machine learning is one of the new technology trends. In this digital world, machine learning, imaging, classification, analysis, regression, etc. It is increasingly used in many fields and industries.

It is possible to create and research automation without special instructions. There are three types of machine learning: supervised machine learning, unsupervised machine learning, and reinforcement machine learning algorithms. In short, machine learning is the science of learning and acting like a human.

Assessing the student's ability is very important and should be well received. In this study, the concept of machine learning was used to determine the next learning level of students. This estimate can be used for any type of school, recruiter, etc. is important for Based on the results of the correct forecast, schools can identify weak employees and provide the necessary training to improve their performance. Employers also spend a lot of money choosing the right candidate

For students, career forecasting can be a daunting task as they

face difficulties in choosing careers that satisfy them as well as their skills and interests. Choosing the right career path can have a huge impact on a student's future professional and personal success.

To make an informed decision, students must consider many factors, including their strengths, passions, values, and career prospects. They will also need to research a variety of job and career opportunities, conduct interviews and seek guidance from a mentor or mentor. In such situations, predictive models and tools can provide valuable information to help students evaluate career options and make more informed decisions. The model uses a data-driven approach to identify patterns and trends in the market and match skills and personal preferences to career opportunities.

we will explore some of the key factors affecting student performance and discuss the benefits and limitations of predictive models in this context. We will also provide some tips and ideas to help students make better career decisions.

The results of the prediction model are also used to find out whether students are willing to study or want to study their higher education. This research project is mainly focused on estimating the work of undergraduate students.

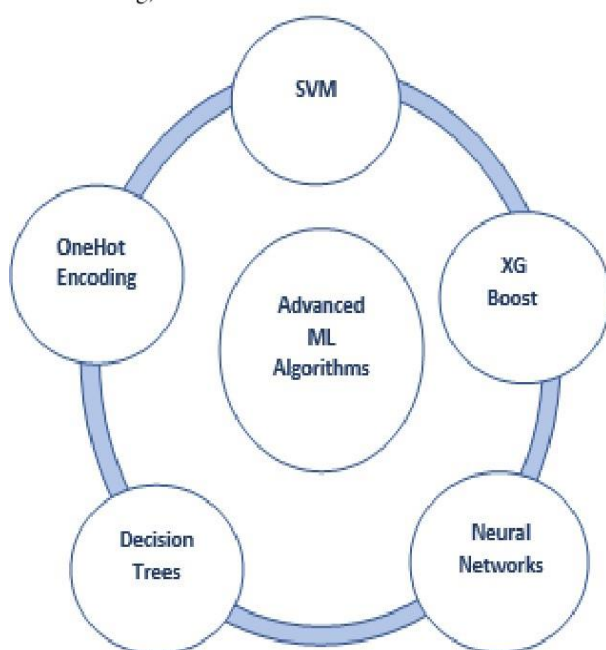


Figure 1 Advance machine learning algorithm [1]

Machine learning algorithms such as SVM, DT, RF and Adaboost classifiers are used to build the model. Among the above products, RF performs better. These classifiers are used with the help of the python programming language because

most of the time problems are easily implemented by this programming language. The next section presents the theories and methods used by various authors in forecasting research.

Literature Review

1. Iqbal et al discuss various machine learning techniques for predicting student performance on a variety of subjects. Methods such as matrix factorization, classification and regression were used in the analysis of data collected by ITU Pakistan. They evaluated performance using machine learning methods and found that RPM is the best among many machine learning methods (Iqbal, 2017).
2. Vaidu et al. used machine learning based on student performance to predict job skills. They used KNN and Naive Bayes model to divide students into different groups. The KNN algorithm predicts the student function with an accuracy of 95.33%, which is the purpose of Naive Bayes67.
3. Predicting our future performance after students, Byung-Hak et al. using a deep learning-based Grit Net algorithm. According to this research paper, Grit Net gives more accurate results based on logistic regression. They received information from Udacity Nanodegree Programs (Schaar, 2017).
4. Jie et al. also proposed a machine learning method to predict student performance in higher education programs. In this study, students' past and present performances are evaluated. It uses a two-tier model with multiple levels of prediction and a data-driven approach for bottom-up prediction based on quality results. Research data show that the proposed method gives better results than the basic method.
5. Pojon Murat reviews machine learning algorithms used to predict student performance. Pojon Murat used three different algorithms: Linear Regression, Naive Bayesian Classification and Decision Trees on two different datasets, of which Roberson et al. is a prominent engineering version. According to the results, Naive Bayes is the best method since it gives 98% accuracy for the first data, while Decision Tree is the best method because it gives 78% accuracy for the second data.

6. Singh et al. They used some machine learning techniques to predict students' performance in engineering courses. By examining the grades of the courses in the previous semester, they predict the grades of the students in the next courses. For this, decision tree classification and Naive Bayes method were used and it was seen that the decision tree gave more accurate results than Naive Bayes (Bendengnu Ksung, 2018).
7. Using machine learning like support vector machines, random forests, gradient boosting and Naive Bayes, Pushpa S and others. It can predict whether a student will fail or pass a grade in the previous semester. As for the prediction, Random Forest has an accuracy of 89.06%, which is more than other algorithms (Pushpa, 2017).
8. Bendengnu Ksung et al. Using DNN Models, eg. Deep Neural Networks for Predicting Student Webology (ISSN: 1735-188X) Vol. 18, no. 6, 2021 3509 <http://www.webology.org> reports. A research paper by Bendengnu Ksung et al. compared DNNs to machine learning algorithms such as Naive Bayes, Neural Networks, and Decision Trees. Accordingly, DNN achieved an accuracy of 84.3%, which is the performance of machine learning (Gerritsen, 2017).
9. To predict student performance, Gerritsen L. et al. using data from learning management about data learned from the use of neural networks. For this article, it is accepted that the Moodle log dataset contains the data of 4601 students. In this paper, the performance of the neural network is compared with six known as K-Nearest Neighbors, Naive Bayes, Decision Trees, Support Vector Machines, Logistic Regression and Random Forests. According to the data, the neural network gives more accurate results than the other six groups (Hernandez, 2018).
10. To estimate a school's exit list, Martin S. et al. It uses four machine learning methods to measure performance. The four methods include random forest, support vector machine, logistic regression, and neural networks. For this research data, data from students from the Costa Rican Institute of Technology (ITCR) between 2011 and 2016 were used. Among the four algorithms, the random forest algorithm is the best choice for college education cost estimations.
11. K. Sripath Roy (2018) Using machine learning algorithms such as SVM decision tree and XG boost to build a student work prediction model. Among these algorithms, SVM reaches 90.3% and gives more accurate results.
12. Mubarak Al Baraka Umar Jesus uses neural network technology to predict student learning. In this study, a neural network model was developed to predict a student's GPA using the student's personal information, opposition, and academic records. According to this model, the prediction accuracy is 73.68%
13. Pardos, Baker, San Pedro, Gowda, and Gowda (2014) also studied Service, but focused on discussing students' thinking and collaborative behavior. Betting math tests can help students get bored, focused, confused, etc. They developed an intervention and collaborative behavior system to measure outcomes using eight technologies to learn to predict outcomes. They also developed a model to predict students' math test scores and showed that the developed tool helped the model achieve high-level predictions.
14. Baker, Berning, Gowda, Zhang, and Hawn (2019) present a case study that uses data on student discipline, attendance, academics, and grades to identify students at high risk of failing in high school. The regression model not only helps authors select student at-risk

Comparative Analysis

Table 1 summarizes the research into several Student analysis approaches presented by various researchers.

Table 1. Comparative Analysis of already existing DR techniques

Author	Method Used	Dataset	Accuracy	Remarks
Pojon Murat (2013)	Linear regression, Naïve Bayes classification, and decision tree	student dataset	For first 98% for second 78%.	Found that Naïve Bayes is the best technique for the first dataset with an accuracy of 98%, while Decision Tree is the best technique for the second dataset with an accuracy of 78%.
(Zimmermann et al., 2015)	Regression models	171 students data from ETH Zurich	N.A	Third year GPA of undergraduate can predict graduate performance.
(Sweeney et al., 2015)	Factorization Machine	33000 GMU students data of fall 2014.	N.A	FM model can predict performance with lower prediction error.
(Elbadrawy et al., 2016) P	Personalized MultiLinear Regression models (PLMR)	30,754 GMU, 14,505 UMN and 13,	N.A	PLMR and MF can predict next term grades with lower error.
(Meier et al., 2016)	Regression and classification	700 UCLA undergraduate students data	N.A	In-class evaluations enables timely identification of weak students.
Vaidu et al 2017 [2]	KNN and Naïve Bayes	Student dataset	95.33% and 67.67%	Used predict the employability skills of students based on their performance
Pushpa S et al (2017)	Support Vector Machine, Random Forest, Gradient Boosting, and Naïve Bayes algorithms.	student dataset	89.06%.	They found that Random Forest algorithm is more accurate
Bendengnu Ksung et al (2017)	Naïve Bayes, ANN, and Decision Tree algorithms	student dataset	84.3%	Deep Neural Network to predict student Webology performance
(Xu et al., 2017)	Latent factor method based on course clustering	1169 UCLA undergraduate students data	N.A.	Latent factor method performs better than benchmark approaches.
Gerritsen L. et al (2018)	K-Nearest Neighbor, Naïve Bayes, Decision Tree, Support Vector Machine, Logistic	4601 students dataset	Not specified	Neural Network is more accurate compared to the other six classifiers.

	Regression, and Random Forest			
K. Sripath Roy (2018)	SVM decision tree and XG boost	Student dataset	90.3%.	Among these algorithms, the Support Vector Machine gives the more accurate
Bendengnu Ksung, 2018	Support Vector Machine K-Nearest Neighbor Neural Network	third semester performance of MCA students	98%	Neural Network has the highest prediction accuracy by (98%)
Mubarak Al Baraka Umar Jesus	artificial neural network	student's GPA	73.68%	This algo using their personal information, place of resistance, and academic information.
Baker, Berning, Gowda, Zhang, and Hawn (2019	logistic regression classification algorithms	4,864 students	75%	Student who would drop out from a student who would not 76% of the time forentirely new student

Proposed Framework

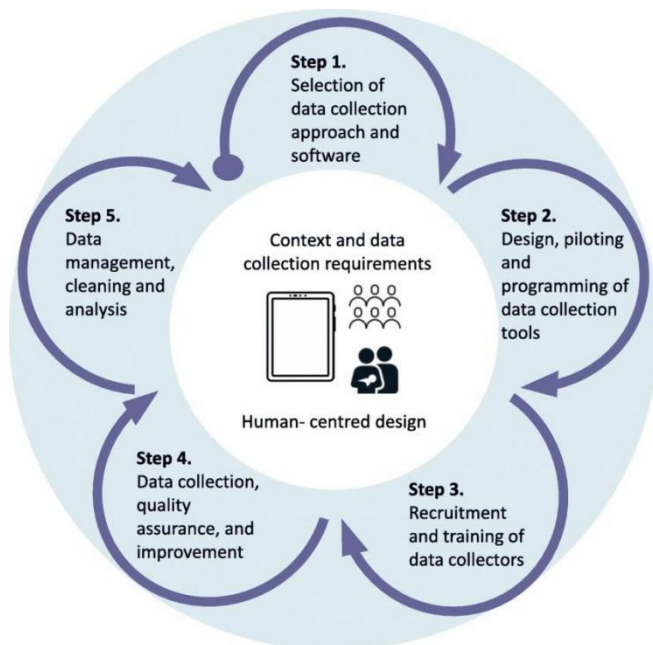


Figure 1 Data Collection Ways[31]

Data collection is one of the most important tasks of any machine learning. Because the input we give to machine learning algorithms is data. Therefore, the efficiency and accuracy of the algorithm depends on the accuracy and quality of the collected data. So the correct information will be output. In order for students to predict the job, there must be many negative factors such as students' performance in various subjects, learning, programming and analytical skills, memory, relationships, interests, personal content such as competitions, hackathons, games, seminars, certificates. interest books etc.) aspects of results.

All these factors are taken into account as they play an important role in determining students' progress in study. Data is collected in different ways. Some of the information was collected by the employees working in different organizations, some of the information was produced, and some of it was obtained from the student records of the school.

Data Pre-processing :-

Gathering information is one job, making it useful is another important job. Information gathered from various sources may be in a consolidated form and may contain a lot of unnecessary, inaccurate and undesirable information.

Furniture cleaning is one of the most important tasks of any machine learning. Because the input we give to machine learning algorithms is data. Therefore, the efficiency and accuracy of the algorithm depends on the accuracy and quality of the collected data. Therefore, the information displayed will be as accurate as possible. In order for students to predict the job, there must be many negative factors such as students' performance in various subjects, learning, programming and analytical skills, memory, relationships, interests, personal content such as competitions, hackathons, games, seminars, certificates. interest books etc.) aspects of results. All these factors are taken into account as they play an important role in determining students' progress in study. Data is collected in different ways. Some of the information was collected by the staff working in different institutions, some of it was produced, and other information came from the student records of the school. In total, about 15,000 documents and 35 articles were collected. Gathering information is one job, making it useful is another important job.



Figure 2 Data preprocessing [32]

Data collected from various sources will be in an unorganized format and there may be improper data and replacing them with appropriate or approximate data and removing null and missing data and replacing them with some fixed alternate values are the basic steps in pre-processing of data. Even data collected may completely contain garbage values. It may not be in an exact format or way which is meant to be.

All such data must be verified and replaced with alternate values to make data meaningful and useful for further processing. Data must be kept in an organized format.

OneHot Encoding :-

OneHot Coding is a process in which categorical values found in data are converted into numbers or other types of patterns to feed into machine learning algorithms and yield better prediction results. OneHot encoding converts categorical values into optimal form as input to various machine learning algorithms. This algorithm is relevant to almost all machine learning algorithms. A few algorithms, such as Random Forest, can handle values with accuracy. In this case, no coding is required.

The OneHot encoding process may seem complicated, but most modern machine learning algorithms will solve this problem. The process here is easy to explain: For example, if a dataset has values like yes and no, an integer encoder gives them values like 1 and 0. Keep setting values from 1 for yes to 0 for no. As long as we assign or assign an integer to this particular text, it's called integer encoding. But here the relationship is important because if we then reverse the encoding, we need to decode the results, especially on the estimated data.

Some machine learning algorithms can work directly with categorical data depending on implementation, such as a decision tree, but most require any inputs or outputs variables to be a number, or numeric in value. This means that any categorical data must be mapped to integers.

The next step is to create a vector for each integer value. Suppose this vector is binary and the length of two possible integers is 2.

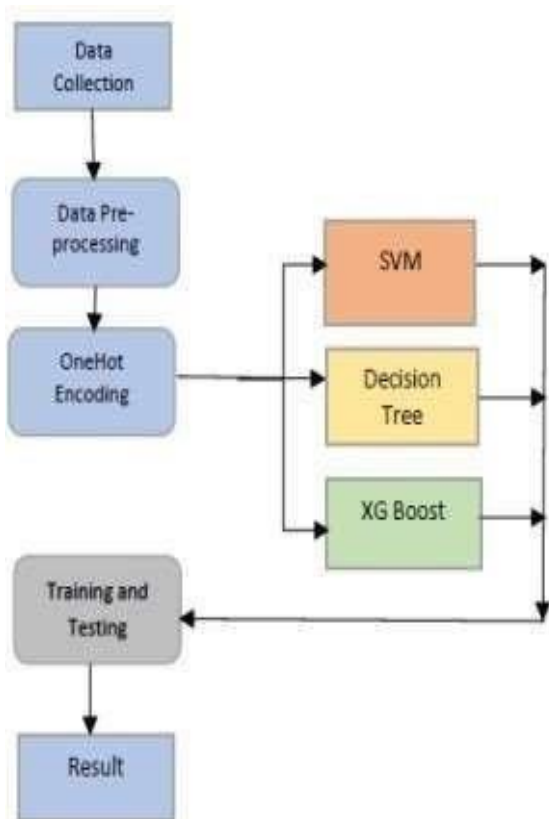


Figure 3 proposed framework hierarchy [33]

Challenges and Limitations

1. **Rapidly changing job market:** The job market is continuously evolving, and new careers are emerging regularly. Therefore, predicting the career path of students in such a dynamic environment is a challenge.
2. **Bias and stereotypes:** The career prediction process can be influenced by bias and stereotypes. For example, students may be encouraged to choose certain career paths based on their gender, race, or socioeconomic status. Therefore, it is essential to eliminate bias and stereotypes from the process.
3. **Lack of career counseling:** Many students do not have access to career counseling, which can help them make informed decisions about their career path. Therefore, providing career counseling services to all students is necessary to ensure that they make the right choices.

Conclusion

Machine learning algorithms have provided a promising solution for student career prediction. The review highlights the different factors affecting career prediction,

the machine learning algorithms used for career prediction, and the limitations of current research. Future research should focus on developing methods to overcome these limitations and provide accurate and interpretable career predictions. Educational institutions and industries can utilize these predictions to provide personalized career guidance and improve student employability

REFERENCES

- [1] P.Kavipriya, "A Review On Predicting Students' Academic Performance Earlier, Using Data Mining Techniques", International Journal Of Advanced Research In Computer Science And Software Engineering.
- [2] Ali Daud, Naif Radi Aljohani, "Predicting Student Performance Using Advanced Learning Analytics", 2017 International World Wide Web Conference Committee (Iw3c2).
- [3] Marium-E-Jannat, Saymasultana, Munira Akther, "A Probabilistic Machine Learning Approach For Eligible Candidate Se-lection", International Journal Of Computer Applications (0975 – 8887) Volume 144 – No.10, June 2016.
- [4] Sudheep Elayidom, Dr.Sumam Mary Idikkula, "Applying Data Mining Using Statistical Techniques For Career Selec-Tion", International Journal Of Recent Trends In Engineering, Vol. 1, No. 1, May 2009.
- [5] Dr.Mahendra Tiwari, Manmohan Mishra, "Accuracy Estimation Of Classification Algorithms With Demp Model", International Journal Of © 2022 Jetir April 2022, Volume 9, Issue 4 Wwww.Jetir.Org (Issn-2349-5162) Jetir2204639 Journal Of Emerging Technologies And Innovative Research (Jetir) Wwww.Jetir.Org G268 Advanced Research In Computer And Communication Engineering Vol. 2, Issue 11, November 2013.
- [6] Ms.Roshani Ade, Dr. P. R. Deshmukh, "An Incremental Ensemble Of Classifiers As A Technique For Prediction Of Student's Career Choice", 2014 First International Conference On Networks & Soft Computing.
- [7] Nikita Gorad, Ishani Zalte, "Career Counselling Using Data Mining", International Journal Of Innovative Research In Computer And Communication Engineering.
- [8] Bo Guo, Rui Zhang, "Predicting Students Performance In Educational Data Mining", 2015 International Symposium On Educational Technology
- [9] Ali Daud , Naif Radi Aljohani , "Predicting Student Performance Using Advanced Learning Analytics" .
- [10] Rutvija Pandya Jayati Pandya, "C5.0 Algorithm To

Improved Decision Tree With Feature Selection And Reduced Error Pruning“, International Journal Of Computer Applications (0975 – 8887) Volume 117 – No. 16, May

[11] Ade R. and Deshmukh P. R. (2014). Classification of Students Using psychometric tests with the help of Incremental Naive Bayes Algorithm. International Journal of Computer Applications. (0975 – 8887) Volume 89 – No 14.

[12]. Subahi A., F. (2018). Data Collection for Career Path Prediction Based on Analyzing Body of Knowledge of Computer Science Degrees. Journal of Software. Volume 13.

[13]. Liu Y., Zhang L., Nie L., Yan Y., Rosenblum D. S. (2016). Fortune Teller: Predicting Your Career Path. Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16).

[14]. Uhler B. D., Hurn J. E. (2013). Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective. Journal of Interactive Online Learning. Volume 12.

[15]. Nie M., Xiong Z., Zhong R., Deng W., Yang G. (2020). Career Choice Prediction Based on Campus Big Data—Mining the Potential Behavior of College Students. Applied science. a. Doi: 10.3390/app10082841.

[16]. Badarenah A. A., Alsakran J. (2016). An Automated Recommender System for Course Selection. International Journal of Advanced Computer Science and Applications, Vol. 7, No. 3.

[17]. Gorad N., Zalte I., Nandi A., Nayak D. (2017). Career Counselling Using Data Mining. International Journal of Innovative Research in Computer and Communication Engineering. Vol. 5, Issue 4.

[18]. Chaudhary D., Prajapati H., Rathod R., Patel P., Gurjwar R. K. (2019). Student Future Prediction Using Machine Learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. Volume 5, Issue 2.

[19]. Mourya V., Phatale S., Thakur S., Mane P. (2020). Career Guide Application using ML. International Research Journal of Engineering and Technology (IRJET). Volume: 07 Issue: 09.

[20]. Prasanna L., Haritha D. (2019). Smart Career Guidance and Recommendation System. International Journal of Engineering Development and Research.

Volume7, Issue 3.

[21]. Roy K. S., Roopkanth K., Uday V., Bhavana V., Priyanka J. (2018). Student Career Prediction Using Advanced Machine Learning Techniques. International Journal of Engineering & Technology.

[22]. Umar M. A. (2019). Student Academic Performance Prediction using Artificial Neural Networks: A Case Study. International Journal of Computer Applications (0975 – 8887) Volume 178.

[23]. Ezenkwu C.P., Johnson E.H., Jerome O.B. (2017). Automated Career Guidance Expert System Using Case-Based Reasoning Technique. Cisd ijournal. Volume 8, No.1.

[24]. Amieh E. A., Hamtini T., Aljarah I. (2016). Mining Educational Data to Predict Student's Academic Performance using Ensemble Methods. International Journal of Database

[25]. Elayidom S., Idikkula S. M., Alexander J. (2009). Applying Data mining using Statistical Techniques for Career Selection. International Journal of Recent Trends in Engineering, Vol. 1, No. 1.

[26]. Nawaz M., Adnan A., Tariq U., Salman J. F., Asjad R., Tamoor M. (2014). Automated Career Counseling System for Students using CBR and J48. Journal of Applied Environmental and Biological Sciences.

[27]. Anthony V., Naidoo. (1998). Career Maturity: A Review of Four Decades of Research. Educational Resources Information Centre (ERIC).

[28]. Buhlmann P. (2012). Bagging, Boosting and Ensemble Methods. ETH Zurich, Seminar fur Statistik, HG G17, CH-8092 Zurich, Switzerland.

[29]. Miškovic V. (2014). Machine Learning of Hybrid Classification Models for Decision Support. SINTEZA. [30].

Vafeiadis T., Diamantaras K.I., Sarigiannidis G., Chatzisavvas K.CH. (2015). A comparison of machine learning techniques for customer churn prediction. Elsevier.

[31] https://www.researchgate.net/figure/Five-step-framework-for-data-collection-and-implementation-of-a-large-multi-country_fig1_350404920

[32] https://assets-global.website-files.com/5d7b77b063a9066d83e1209c/6137488ee132f03deba81f9f_data-preprocessing.png

[33] <https://www.semanticscholar.org/paper/Student-Career-Prediction-Using-Advanced-Machine-Roy-Roopkanth/8e580d2c5aa67ed26dfd4b0d8270ca666b749d8f>