International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04923
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Customer Churn Prediction using Data Science

B KUMARI, TUMMALA VINAYA PRASAD
Assistant professor, MCA Final Semester, Master of Computer Applications, Sanketika Vidya Parishad Engineering College, Vishakhapatnam, Andhra Pradesh, India.

ABSTRACT:

The project deals with building a machine learning model to find customer churn. It is used in various kinds of marketing strategy. Customer churn is a big problem and it is one of the important concern for large tech companies due to the straight effect on the revenues of those companies. Due to increase in the use of internet large e-commerce and online selling business, it will be most important to find an unlike customer. A company will not spend the advertisement budget target to individual who are unlikely to comeback instead of finding who are coming back and buying products, so the company will be improving the sales and profit. Here we develop machine learning models to predict and identify those loyal customers. Hence, we can spend advertisements only on particular customers to improve company growth in the future. It is mostly useful in subscription based business.

Key words: Data Preprocessing, Data Analysis, Predict customer churn.

1. INTRODUCTION:

- Churn Prediction : It is one of the most important big data use case in business. It detects the customers who are likely to cancel the subscription to a service or move out of company.
- Churn : It is nothing but losing the customers. Is a problem for service based companies because it is more expensive to acquire a new customer than to keep the existing one.

PROJECT OBJECTIVE:

- Data Preprocessing.
- Data Analysis.
- Use various Machine Learning and Deep Learning Algorithms to find the prediction models, evaluate, calculate accuracy and performance of the models.
- Finding the best model for our business case.
- Predict customer churn

LITERATURE SURVEY:

- Anuj Sharma et al.[1] Customer Churn in Cellular Network Services. Used model ANN, accuracy 92% ,dataset from uci Repository.
- Abdelrahim Kasem Ahmad et al. [2] Customer Churn in Cellular Network Services. The data set is size 70 tb for processing they used hadoop. Used model is XGBoost with AUC score 9.3%.

- Eunio lee et al. [3] online gamming (MMORPGs) based on korean company, data is 10 lakhs of users and model is XGboost accuracy is 90%.
- Adnan Amin at al. [4] Cross-Company Churn Prediction (CCCP) is a domain of research where one company (target) is lacking enough data and can use data from another company (source) to predict customer churn successfully model used Single Rule Induction with AUC 0.541.
- Lomax, S, and Vadera, S et al. [5] CRM(customer relationship management) Analysis of customers, the data set used for own business privacy but dataset contains 10 attributes model used is Decision Tree with accuracy 90%.
- Aditya Kulkarni at al. [6] Customer churn prediction in telecommunication industry using data certainty the data is grouped into different zones and preprocessed. Used various algorithms and best model is logistic Regression with 80.38% accuracy.

Data set:

- The project is on the business use case of particular Audio company the sell audio books.
- Dataset contain 14062 rows and 12 columns.
- 14062 rows may have some missing values.
- Target is a variable, whether the customer churns or not
- We will develop a model to predict customer churn.

Dataset is imbalanced dataset .

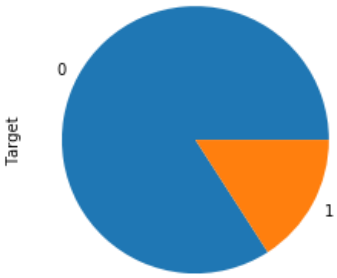
Dataset size is 14062 and columns is 12 included with Target.

0 is not Converted

1 is Converted

0 is 11847 rows of data

1 is 2237 rows of data



EXISTING SYSTEM:

- The advertisements are done digitally, so it actually works based on cookies.
- If the person is visiting a particular company website a unique cookie will be stored in our system and based on that they will start digital advertisement.
- Some companies like Google Ads will charge the advertisement cost.

- The digital marketing will store the cookies of the customer who visits and they provide a unique id for those accounts and they start advertisement on those who visits the sites and that user unique id as target and start advertisement.
- If he visits another site and still the product will be showing in offer which will be a waste of money spending on advertisements those who are not going to buy the products.
- So it is very important to identify the customers that are churning i.e who are not loyal to particular company or service provider.

PROPOSED SYSTEM:

- The proposed system overcomes the problem of the existing system. This project mainly aims at developing the model for predicting customer churn using supervised learning algorithms.
- The model will take the data of an individual customer and analyse . and find that the customer whether he will be buying the product from our company or churn the company
- If he/she is buying our company products it will be good to spend on advertisements to that particular customer, if not we will stop spending on the advertisements to that person.
- With the growth of Machine Learning methods, the proposed system we will analyse the data and we target on those individuals who are better to purchased the product. The model will take the customer's recent activity and its analysis of his previous data and gives the result.
- Machine learning techniques provide better results for prediction by constructing models from datasets with better accuracy.

FUNCTIONAL REQUIREMENTS:

The functionalities includes:

- Select dataset and should be able to analyse and pre-process the data acquired.
- Should be able to train the models.
- Should be able to evaluate accuracy of models.
- Should be able to identify the customer churn using the model.

Software and Hardware Requirements:

- Software Requirements

  Language : Python

  Environment : Jupyter Notebook

  OS : Windows 10

- Hardware Requirements

  RAM : 4GB

  Hard Disk : 40GB

  Monitor : Laptop or Computer

  Processor : Intel i3 or higher

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04923
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

System Design:

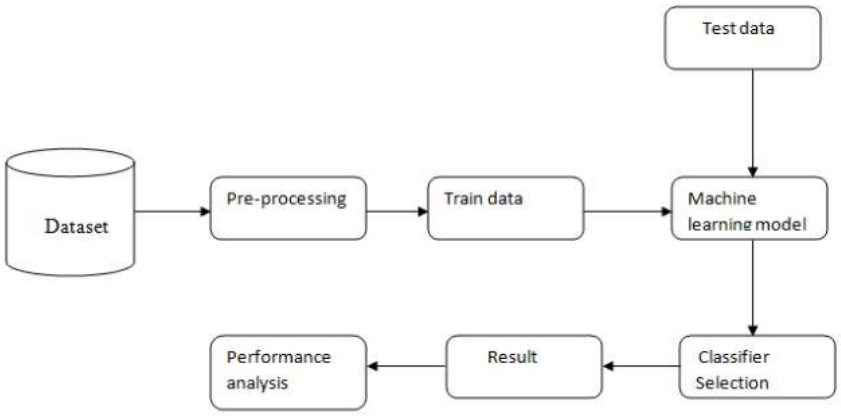
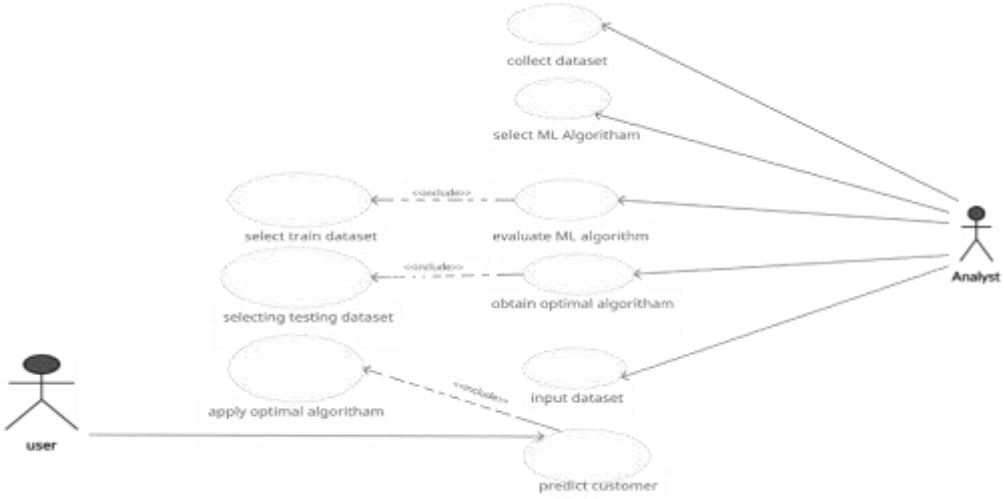


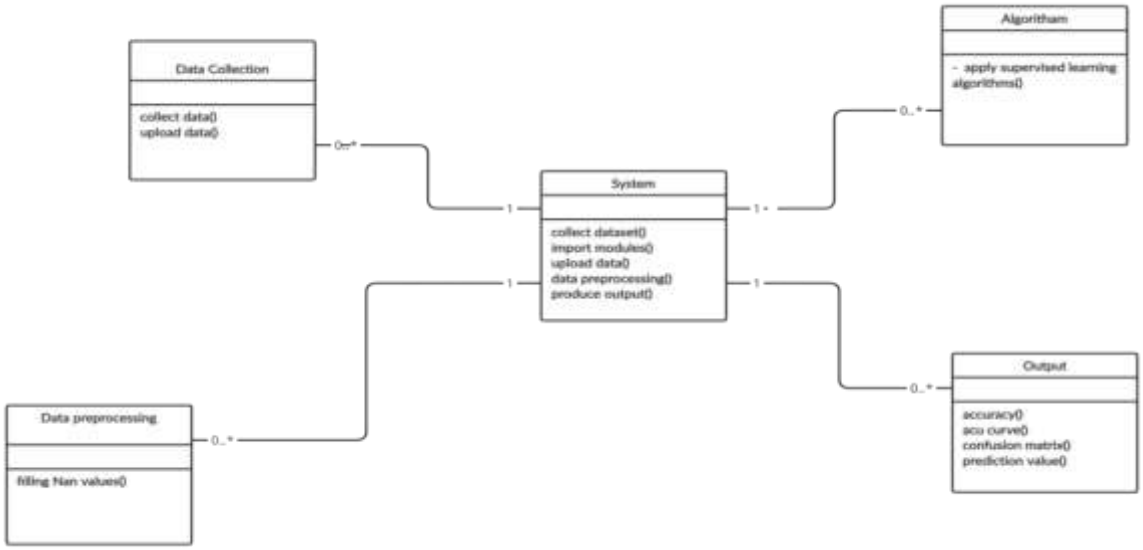
Fig :

UML Design:

- **USECASE DIAGRAM**

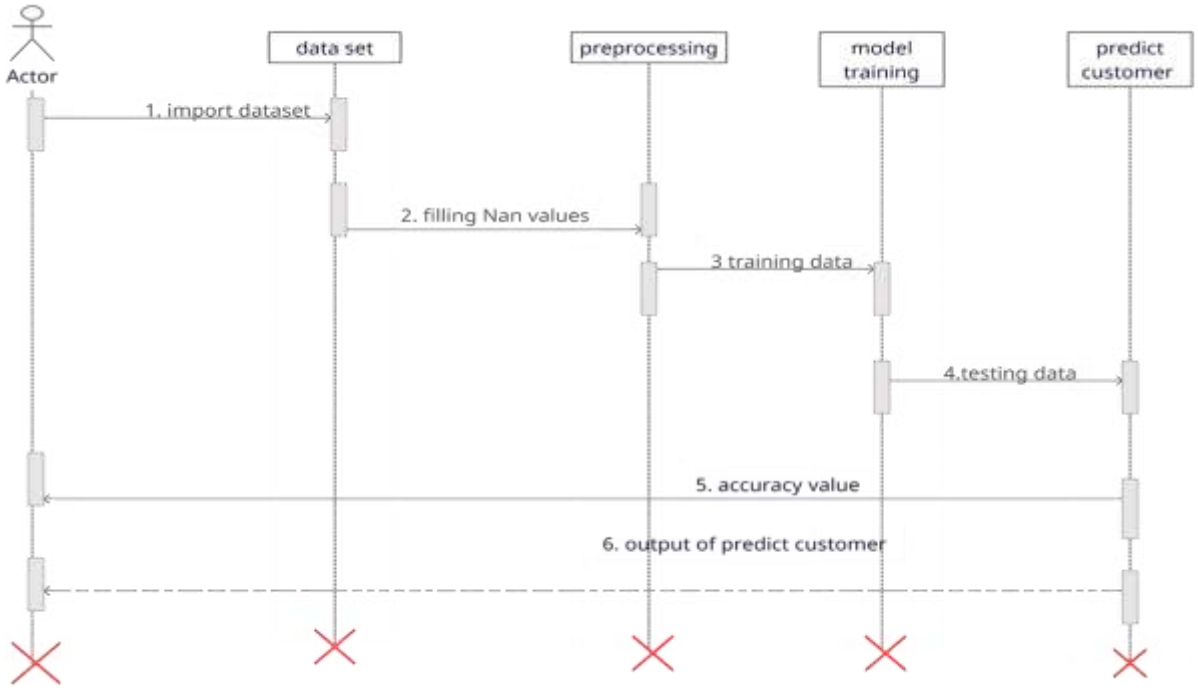


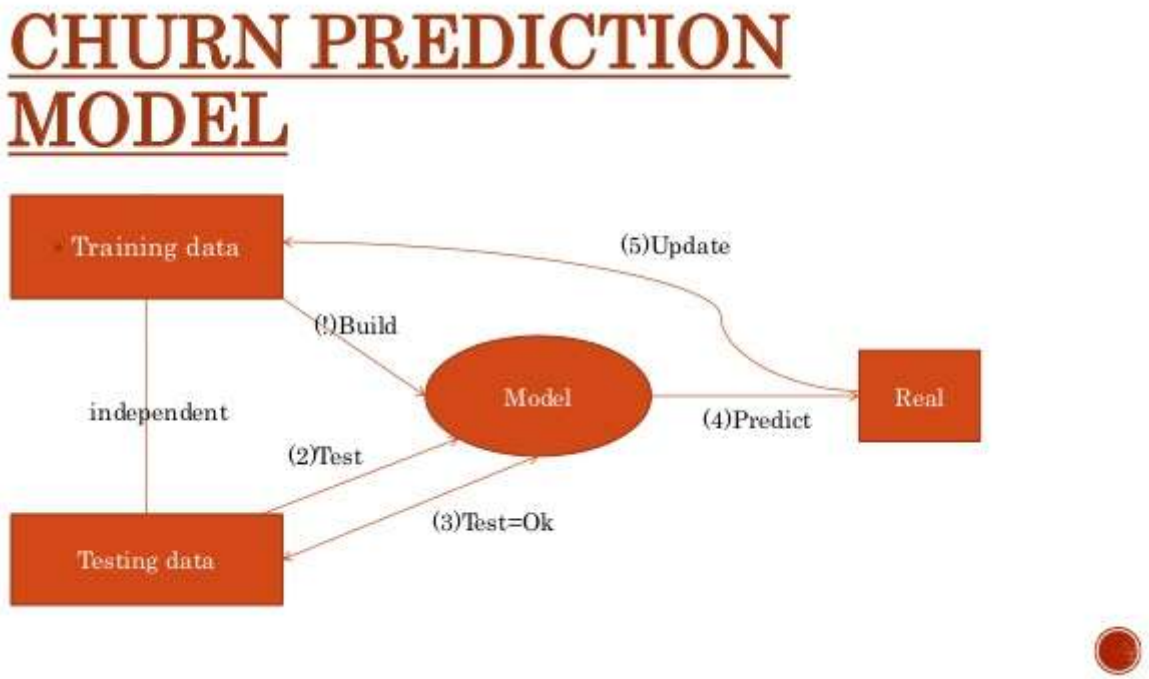
CLASS DIAGRAM:



International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04923
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

SEQUENCE DIAGRAM:





IMPLEMENTATION:

- Import the libraries and load the dataset

Dataset is collected from the Audiobook app, the data consists of a combination of continuous and discrete values. The data set will be in the size of 14062 rows and 12 features including the target. So the dataset which is in .csv format will be collected and will be loaded into the environment.

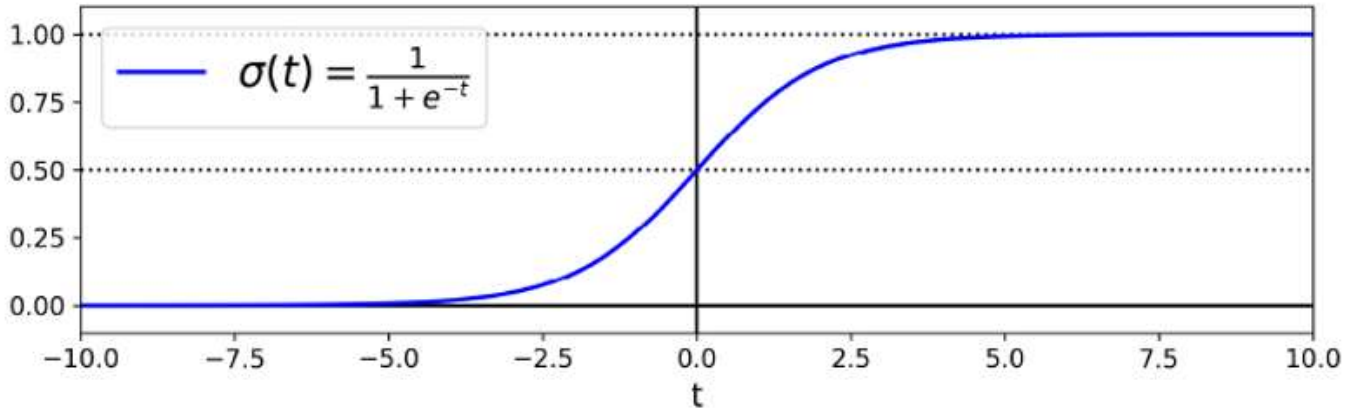
- Preprocess the data We preprocess the data because there may be many missing values or NaN values in the dataset, so after filling missing values we scaling the data and after scaling, normalize the data using normalization techniques which makes the data in good shape and easy to learn the data in the model.

- Data splitting

The dataset is split into 3 segments – training, validation and testing. Training data will be used to build a model..In the data 80% is for training and 10% is for validation and 10% of is for testing.

Logistic Regression:

- Logistic regression - used for performing binary classification problems using probability.
- It calculates the probability that a given value belongs to a specific class.
- We can say that logistic regression also acts as a binary classifier . If the probability is more than 50%, it assigns the value in that particular class else if the probability is less than 50%, the value is assigned to the other class**.**
- Sigmoid Function in logistic Regression:



- Sigmoid function can be used for logistic regression.

- It is a mathematical function that can map real values to any values only in the range 0 and 1. Hence, it is helpful for calculating probability.
- We predict the class based on the value obtained.

$$\phi(z) = \frac{1}{1 + e^{-z}}$$

Here, z is the net input, the linear combination of weights, and the inputs (that is, the features associated with the training examples):

$$z = \boldsymbol{w}^T \boldsymbol{x} = w_0 x_0 + w_1 x_1 + \cdots + w_m x_m$$

- we can conclude that this sigmoid function takes real-number values as input and transforms them into values in the range [0, 1] with an intercept at $\phi\phi(zz) = 0.5$.

- To avoid overfitting and Underfitting we use the Regularization
- Logistic regression is very helpful for binary classification

ISSN: 2583-6129
DOI: 10.55041/ISJEM04923

RESULT ANALYSIS:

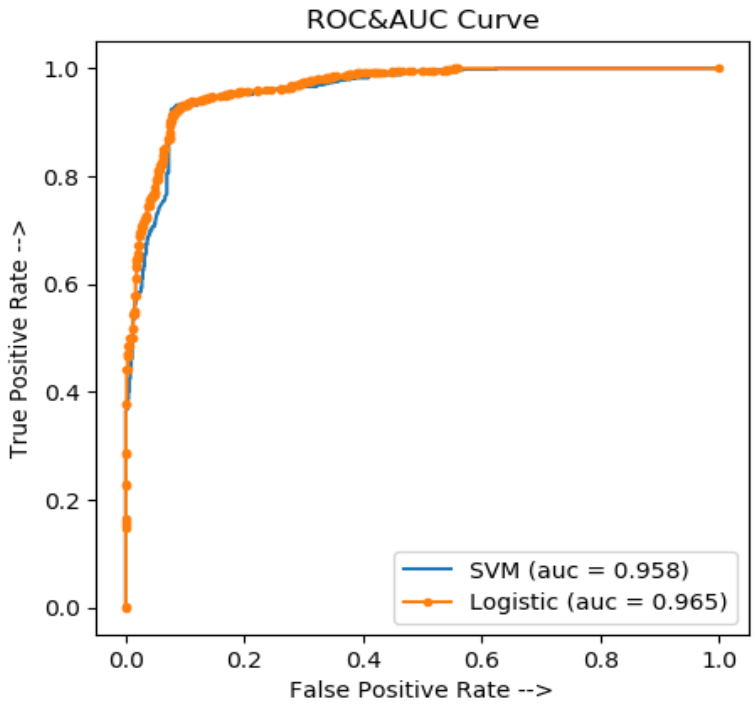
- **Logistic Regression**

• • Model	• Training accuracy	• Testing accuracy	• AUC&ROC curve
• Logistic Regression	• 0.92181682	• 0.9218 •	• 0.965

- **Support Vector Machines**

Model	Training accuracy	Testing accuracy	AUC&ROC curve
SVM	0.9125	0.9263	0.958

ROC & AUC Curve:



CONCLUSION:

- The Customer Churn Prediction is done based on AudioBook dataset. We predicted whether the customer is ready to churn or not. We used logistic regression and svm to train the model. These prediction models need to achieve high AUC values and it is obtained for logistic regression.

- This can further be integrated with the company's software and identify the loyal customers easily.

8. ACKNOWLEDGEMENTS

B.kumari working as a Assistant professor in master of computer application sanketika vidya parishad engineering college, Visakhapatnam Andhra Pradesh. With 2 years of experience in Master of computer applications (MCA), accredited by NAAC.she has a membership in IAENG with her area of intrest in java full stack, DBMS,Data structures and C language.

Tummala vinaya prasad is pursuing his final semester MCA in Sanketika Vidya Parishad Engineering College, accredited with A grade by NAAC, affiliated by Andhra University and approved by AICTE. With interest in Data Science T.vinaya prasad has taken up his PG project on Customer Churn Prediction and published the paper in connection to the project under the guidance of B.Kumari SVPEC.

9. REFERENCES

[1] Investigating customer churn in banking: A machine learning approach and visualization app for data science and management

https://www.sciencedirect.com/science/article/pii/S2666764923000401

[2] A big data analytics model for customer churn prediction in the retiree segment

https://www.sciencedirect.com/science/article/abs/pii/S0268401218301518

[3] Customer churn prediction in telecom using machine learning in big data platform

https://link.springer.com/article/10.1186/s40537-019-0191-6

[4] Customer churn prediction in telecom using big data analytics

https://iopscience.iop.org/article/10.1088/1757-899X/768/5/052070/meta

[5] Customer churn prediction in telecommunication industry using data certainty

https://www.sciencedirect.com/science/article/abs/pii/S0148296318301231

[6] Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior

https://www.emerald.com/jsit/article-abstract/19/1-2/65/251186/Comparison-of-supervised-machine-learning?redirectedFrom=fulltext

[7] Customer churn prediction system: a machine learning approach

https://link.springer.com/article/10.1007/S00607-021-00908-Y

[8] A comparison of machine learning techniques for customer churn prediction

https://www.sciencedirect.com/science/article/abs/pii/S1569190X15000386

[9] The use of knowledge extraction in predicting customer churn in B2B

https://link.springer.com/article/10.1186/s40537-021-00500-3

[10] A novel customer churn prediction model for the telecommunication industry using data transformation methods and feature selection

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0278095

[11] Behavioral attributes and financial churn prediction

https://epjds.epj.org/articles/epjdata/abs/2018/01/13688_2018_Article_165/13688_2018_Article_165.html

[12] Why you should stop predicting customer churn and start using uplift models

https://www.sciencedirect.com/science/article/pii/S0020025519312022

[13] Decoding E-commerce Customer Churn: Harnessing Data Science to Combat Negative Experiences

https://najer.org/najer/article/view/10

[14] Improving customer retention in taxi industry using travel data analytics: A churn prediction study

https://www.sciencedirect.com/science/article/pii/S0969698925000670

[15] Customer churn prediction using improved balanced random forests

https://www.sciencedirect.com/science/article/abs/pii/S0957417408004326

[16] Enhancing customer retention in telecom industry with machine learning driven churn prediction

https://www.nature.com/articles/s41598-024-63750-0

[17] A survey on machine learning methods for churn prediction

https://link.springer.com/article/10.1007/s41060-022-00312-5

[18] Handling class imbalance in customer churn prediction

https://www.sciencedirect.com/science/article/abs/pii/S0957417408002121

[19] A data-driven approach with explainable artificial intelligence for customer churn prediction in the telecommunications industry

https://www.sciencedirect.com/science/article/pii/S2590123025007066

[20] Customer churn prediction in influencer commerce: An application of decision trees

https://www.sciencedirect.com/science/article/pii/S1877050922001703

[21] Forecasting customer churn: Comparing the performance of statistical methods on more than just accuracy

https://journals.open.tudelft.nl/jscms/article/view/6125

[22] Evaluating the Effectiveness of Different Machine Learning Models in Predicting Customer Churn in the USA

https://al-kindipublishers.org/index.php/jbms/article/view/8608

[23] Deep learning for customer churn prediction in e-commerce decision support

https://www.tib-op.org/ojs/index.php/bis/article/view/42

[24] Research on customer churn prediction and model interpretability analysis

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0289724