# CREDIT CARD USAGE PREDICTION

**TEAM 3**

SNEHA DARA - 20137505

SAKETH MADDALA - 20136958

ALEKHYA NUKALA - 20137856

# INTRODUCTION

- As the Technology continues to grow, increasingly customers are using the credit cards to make purchases over the internet.
- As a result, it is crucial for many banks to keep their clients in terms of maintaining their sales and profits.
- Therefore, many banks are starting to appreciate the value of their clients and pay attention to CRM.
- This project uses some of the techniques to detect customer churn for credit cards using estimations.

# DATA COLLECTION

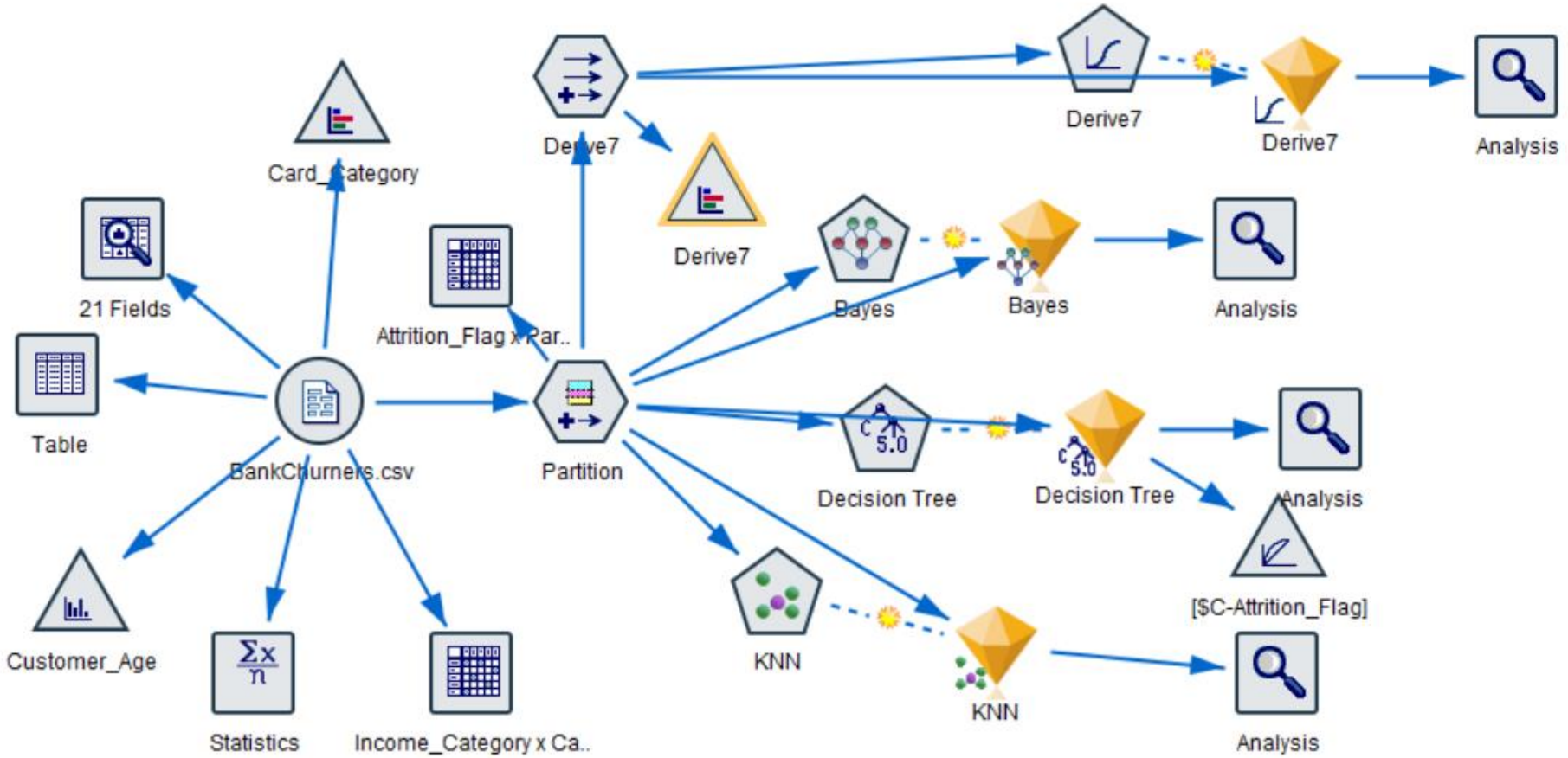
This study was based on a database of bank credit card user loss, which was gathered from kaggle website.

There are more than 9000 clients in this dataset who provide information about their marital status, customer ID, Gender, Education level, Income Category and many more. There are almost 18 characteristics.

The four credit card varieties available to users are silver, gold, blue and platinum. Clients are classified as "churn clients" when they choose to change the bank.

This project uses four techniques namely Bayes, Decision tree, Regression and KNN algorithms.

# SPSS Modeler



# MISSING VALUES

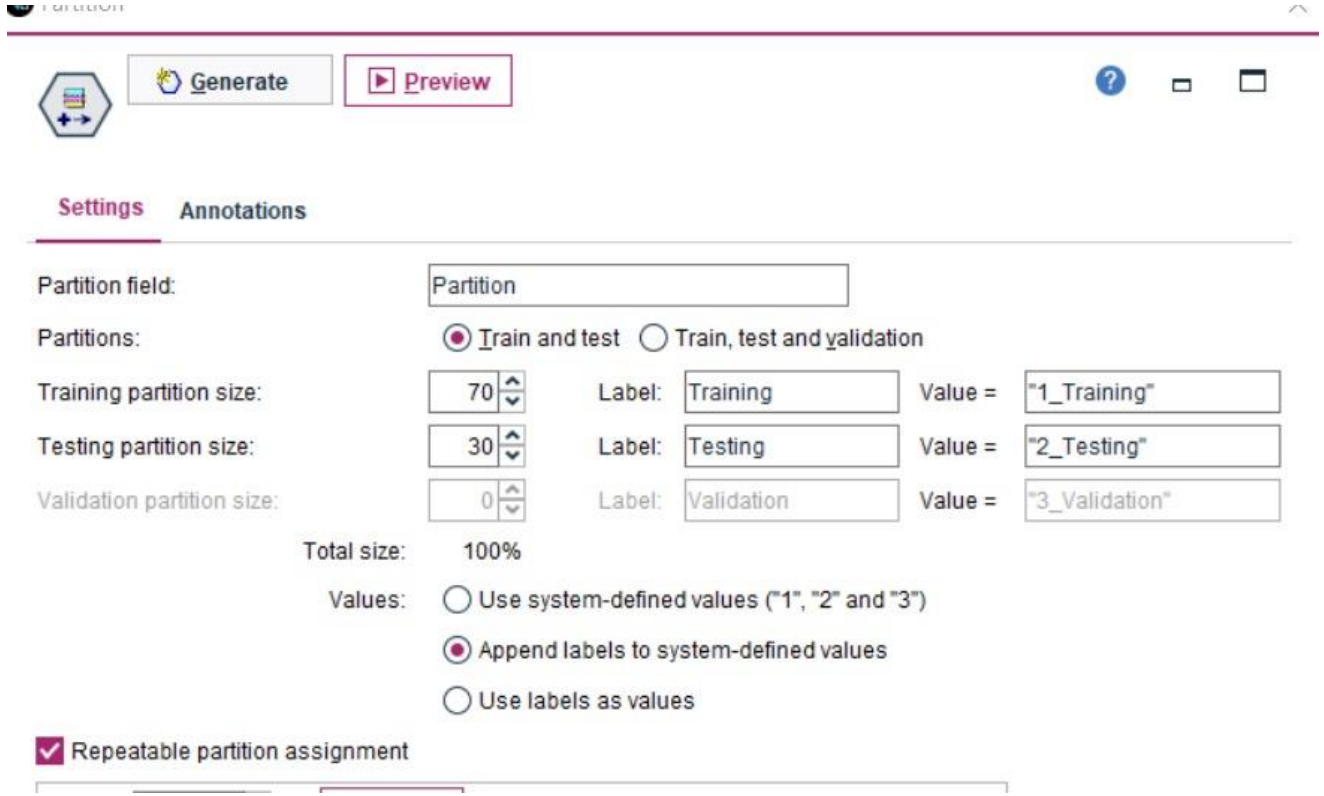
- The values or data that are not collected for one or more fields in the provided dataset are referred to as missing data.
- There are no missing values, as can be seen in the image below.

Complete fields (%): 100% Complete records (%): 100%

| Field | Measurement | Outliers | Extremes | Action | Impute Missing | Method | % Complete | Valid Record |
|---|---|---|---|---|---|---|---|---|
| CLIENTNUM | Continuous | 0 | 0 | None | Never | Fixed | 100 | 10 |
| Attrition_Flag | Flag | -- | -- | -- | Never | Fixed | 100 | 10 |
| Customer_A... | Continuous | 1 | 0 | None | Never | Fixed | 100 | 10 |
| Gender | Flag | -- | -- | -- | Never | Fixed | 100 | 10 |
| Dependent_... | Continuous | 0 | 0 | None | Never | Fixed | 100 | 10 |
| Education_L... | Nominal | -- | -- | -- | Never | Fixed | 100 | 10 |
| Marital_Status | Nominal | -- | -- | -- | Never | Fixed | 100 | 10 |
| Income_Cat... | Nominal | -- | -- | -- | Never | Fixed | 100 | 10 |
| Card_Categ... | Nominal | -- | -- | -- | Never | Fixed | 100 | 10 |
| Months_on_... | Continuous | 0 | 0 | None | Never | Fixed | 100 | 10 |
| Total_Relatio... | Continuous | 0 | 0 | None | Never | Fixed | 100 | 10 |
| Months_Inact... | Continuous | 124 | 0 | None | Never | Fixed | 100 | 10 |
| Contacts_Co... | Continuous | 54 | 0 | None | Never | Fixed | 100 | 10 |
| Credit_Limit | Continuous | 0 | 0 | None | Never | Fixed | 100 | 10 |
| Total_Revolvi... | Continuous | 0 | 0 | None | Never | Fixed | 100 | 10 |
| Avg_Open_T... | Continuous | 0 | 0 | None | Never | Fixed | 100 | 10 |
| Total_Amt_C... | Continuous | 135 | 28 | None | Never | Fixed | 100 | 10 |
| Total_Trans_... | Continuous | 391 | 0 | None | Never | Fixed | 100 | 10 |
| Total_Trans_... | Continuous | 2 | 0 | None | Never | Fixed | 100 | 10 |
| Total_Ct_Ch... | Continuous | 76 | 37 | None | Never | Fixed | 100 | 10 |
| Avg_Utilizatio... | Continuous | 0 | 0 | None | Never | Fixed | 100 | 10 |

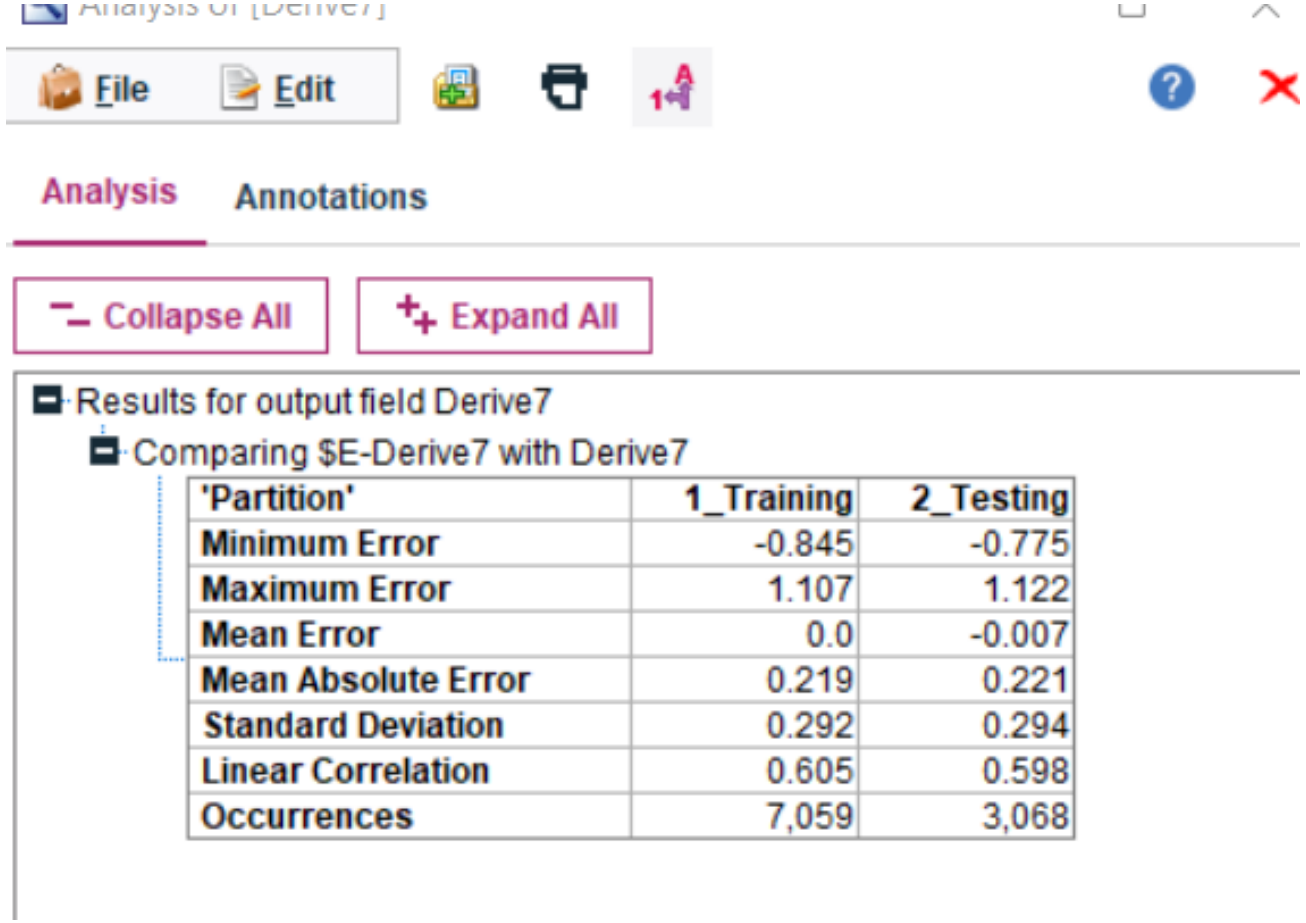
# TRAINING AND TESTING

- A technique to evaluate your model's accuracy is training and testing.
- Before performing the algorithms, the data is sent for training and testing in partition node.
- 70% of training and 30% of testing is performed.



# REGRESSION

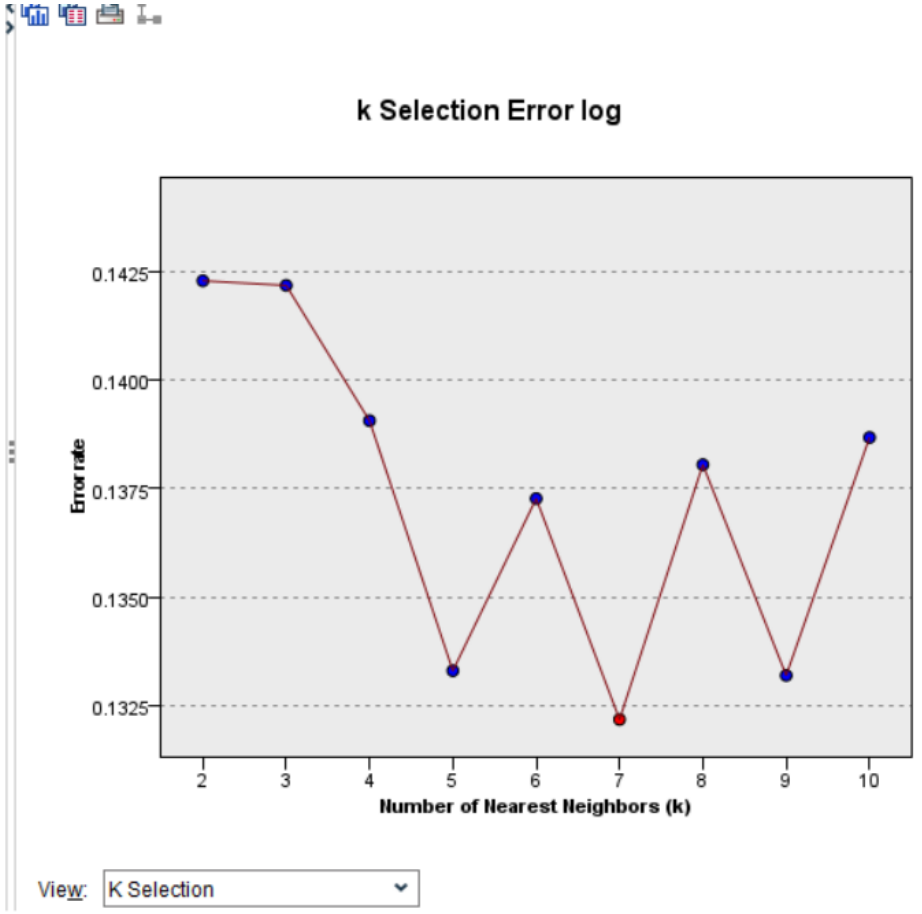
- The continuous-valued variable can be predicted using the supervised method known as regression.




| 'Partition' | 1_Training | 2_Testing |
|---|---|---|
| Minimum Error | -0.845 | -0.775 |
| Maximum Error | 1.107 | 1.122 |
| Mean Error | 0.0 | -0.007 |
| Mean Absolute Error | 0.219 | 0.221 |
| Standard Deviation | 0.292 | 0.294 |
| Linear Correlation | 0.605 | 0.598 |
| Occurrences | 7,059 | 3,068 |

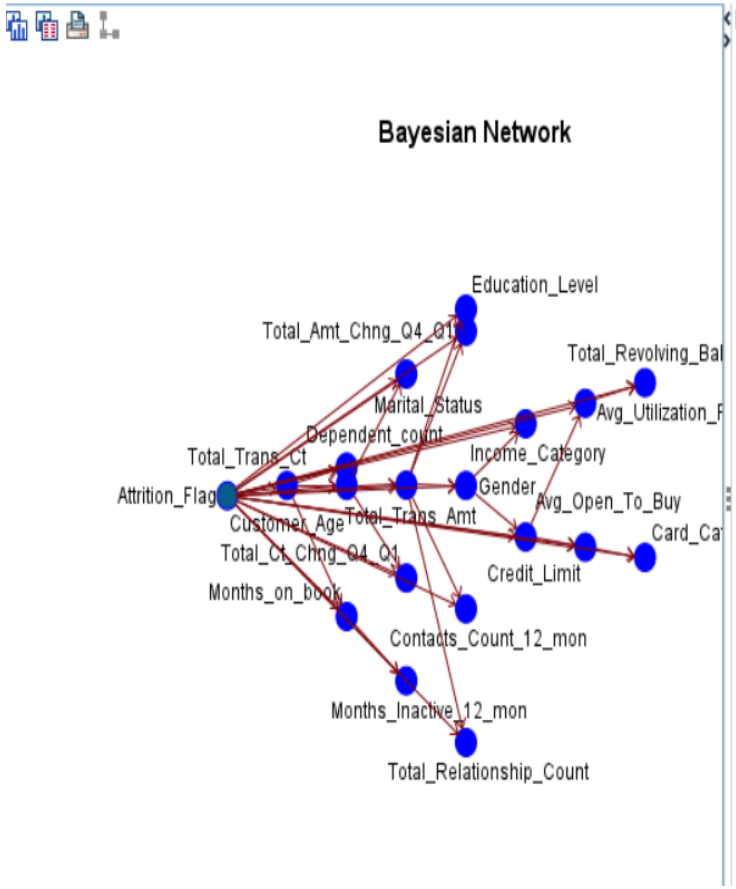
# KNN ALGORITHM

- K nearest neighbor or KNN makes assumptions directly from training set data by determining the k objects that are nearest to the input depending on distance, where k is a parameter that can be changed to affect classifier performance, and it then allocates classification based on the classes that have received the most votes out of these neighboring classes.



# BAYE'S ALGORITHM

- Bayes Algorithm shows the relation between two conditional probabilities which are the reverse of each other.
- Based on the new information Bayes theory revises previously calculated probabilities.

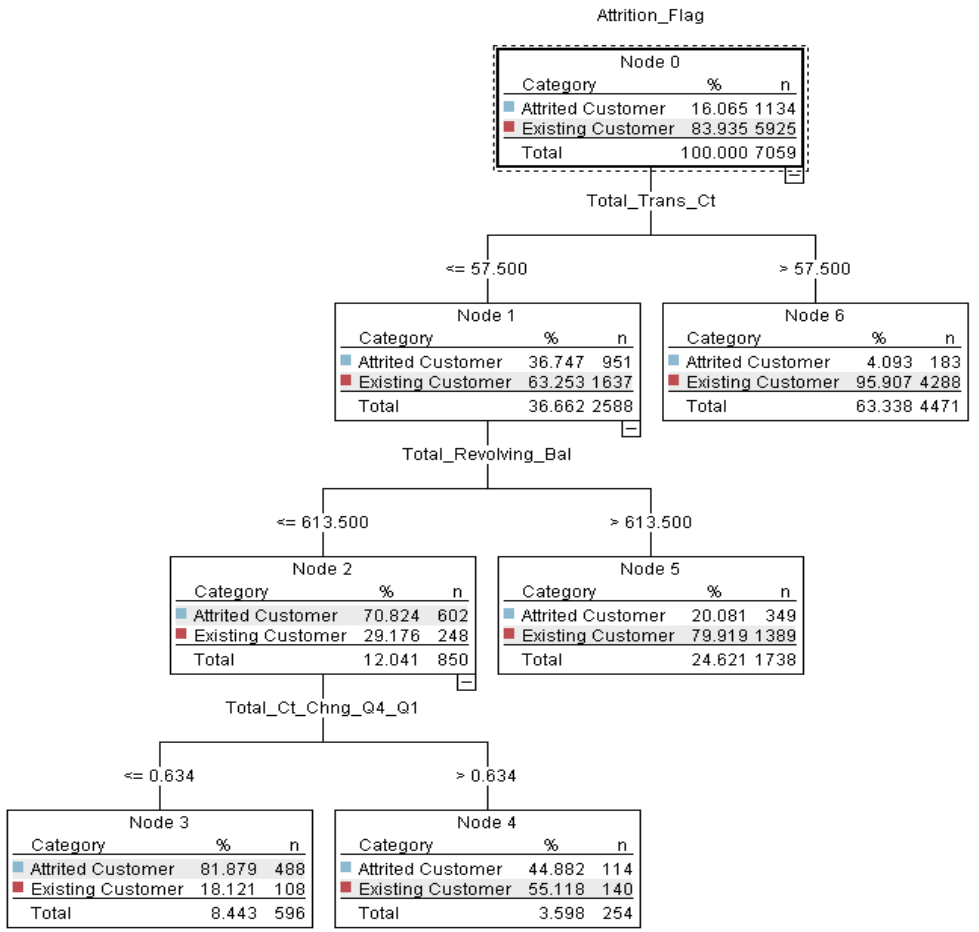



Conditional Probabilities of Attrition_Flag

| Probability | |
|---|---|
| Existing Customer | Attrited Customer |
| 0.84 | 0.16 |

# DECISION TREE

- In the decision tree, the data is continually divided by a specific variable. In this type of learning, you describe what the source is and what the associated output is in the training data.
- Nodes and leaves are the two components that can describe the tree. The results are represented by the leaves. The data is divided in decision nodes.



| | Bayes | KNN | Decision Tree | Logistic Regression |
|---|---|---|---|---|
| **Accuracy** | 0.92 | 0.89 | 0.88 | 0.90 |
| **Recall** | 0.96 | 0.98 | 0.98 | 0.60 |
| **Precision** | 0.94 | 0.89 | 0.89 | 0.74 |
| **FP Rate** | 0.27 | 0.61 | 0.60 | 0.04 |

| | BAYES | KNN | DECISION TREE | LOGISTIC REGRESSION |
|---|---|---|---|---|
| **Build time** | 0.23 | 1.45 | 0.22 | 0.24 |
| **Correct Instances** | 2,852 | 2,731 | 2,729 | 2,771 |
| **Incorrect Instances** | 216 | 337 | 339 | 297 |
| **AUC** | 0.96 | 0.91 | 0.82 | 0.91 |

# FUTURE WORK

- To create a more reliable, more precise, quicker, simpler, and more effective churn prediction model, future research must analyze the best independent variables.

- The optimal outcomes for a particular dataset are not always guaranteed when using the traditional approach. As a result, prediction model accuracy and efficiency should be increased further in future.

# CONCLUSION

- The Aim of the project is to predict the Attired Customer of the credit card company.
- Processed the dataset using the Bayes, Decision Tree and KNN by using the cross validation.
- We found that the last 12 months, total transaction count, last 12 months revolving balance have significant impact.
- More frequent customers use their credit card are less likely to leave.
- Credit card company can increase retention rate which might bring greater profits and by making promotions and offering coupons

# REFERENCES


- The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients (gitbooks.io)
- Credit Card Customer Churn Predictive Analytics | by Trie Sony Kusumowibowo | Oct, 2022 | Dev Genius
- https://www.kaggle.com/datasets/whenamancodes/credit-card-customers-prediction