CUSTOMER CHURN PREDICTION USING MACHINE LEARNING TECHNIQUES

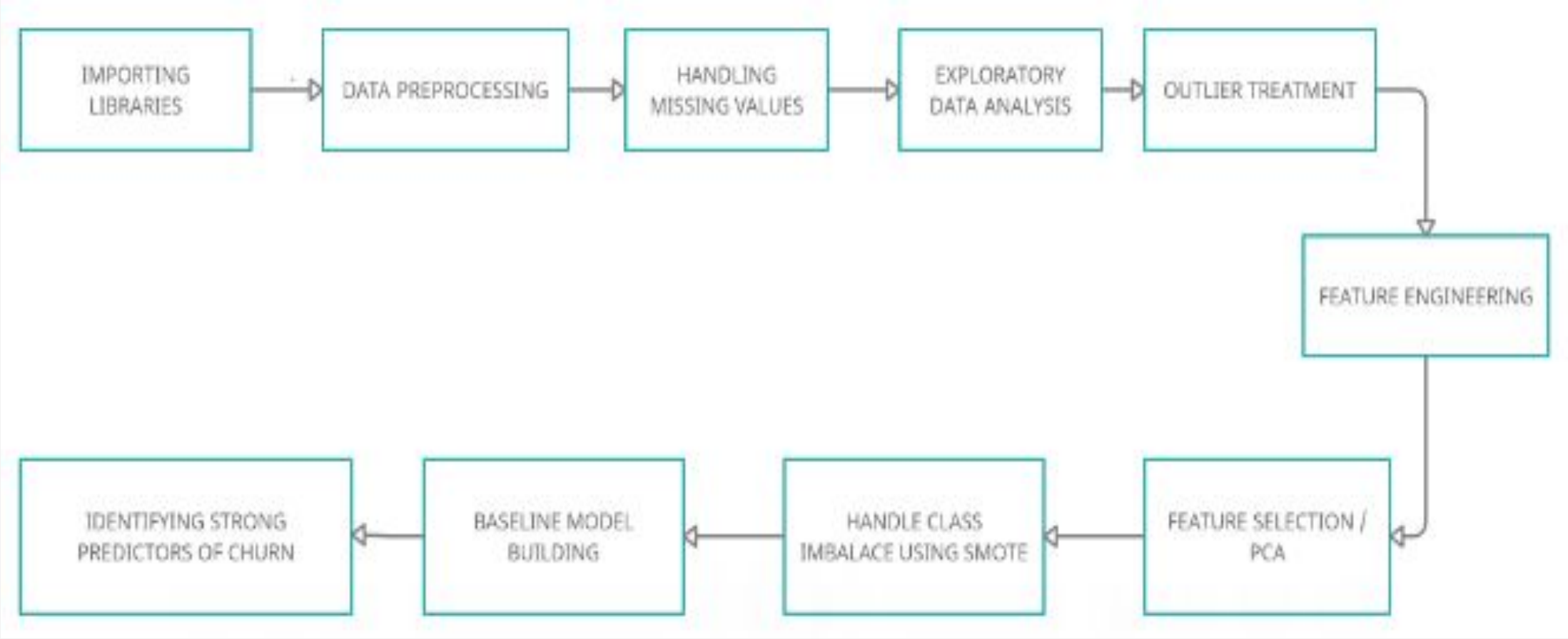
OBJECTIVE

To build a machine learning model that can classify whether the customer who is a part of the organization will still continue to be a part of it or will the customer leave the organization. This type of classification can be used by any sector or industry in the market. It also provides factors behind the churning of churn customers through the rules generated by using the attribute-selected classifier algorithm.

LITERATURE SURVEY

S.NO	PAPER	TECHNIQUE	MERITS	DEMERITS
1.	Predicting Customer Churn Prediction in Telecom Sector Using Various Machine Learning Techniques.	Logistic Regression, SVM, Random Forest, Gradient Boosting Trees.	Better AUC value for Gradient Boosting Trees model.	More data processing and improvement in the AUC vales of the models are required.
2.	Application of customer churn prediction based on weighted selective ensembles.	Bayes, ANN, SVM weighted selective ensembles.	Ensemble methods outperform the individual classifiers.	Difficulty in selecting the parameters of each algorithm to be used as an ensembled classifier.
3.	Customer Churn Prediction for Telecom Services.	Random Forest.	Random Forest with 50 trees and up to maximum depth of 15 gives the best result.	Improvement is required in dealing with dataset impurities.
4.	Churn Prediction using Neural Network based Individual and Ensemble Models.	DL, NN, AUTOMLP, Ensemble classifier Bagging, Majority Voting, AdaBoost.	Bagging with NN outperforms among all the classifiers.	Further research needed on best combination of classifiers to obtain good performance metrics.
5.	A Churn Prediction Model using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector.	Random Forest.	Random Forest produced better F-measure result that is 88%.	Further research required on changing behavior patterns of churn customers by applying Artificial Intelligence techniques

BLOCK DIAGRAM



ALGORITHM

- **Logistic Regression** is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist.
- **Extra tree classifier** is an ensemble learning method fundamentally based on decision trees.
- **K-nearest neighbor (KNN)** algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems
- **AdaBoost,** short for Adaptive Boosting, is a statistical classification meta-algorithm. It can be used in conjunction with many other types of learning algorithms to improve performance.
- **Decision tree** is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.
- **Random forest** is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean/average prediction of the individual trees.

SIMULATION TOOL

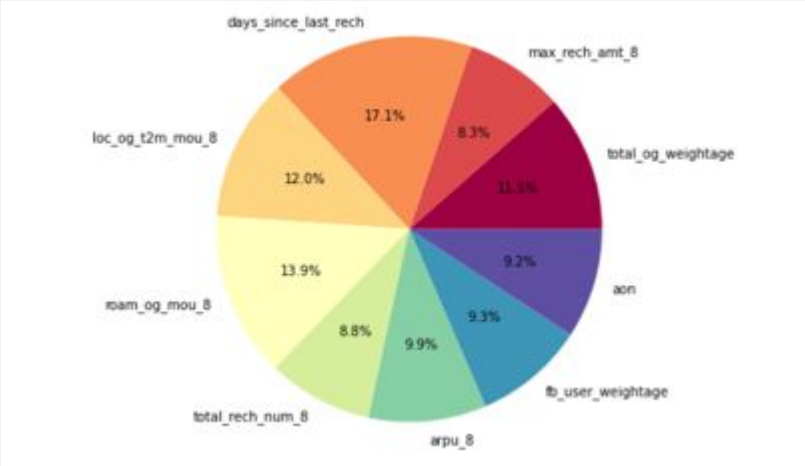
- JUPYTER NOTEBOOK

METHODOLOGY

- DATA PREPARATION
  - Data understanding
  - Data preprocessing
  - Exploratory Data Analysis
  - Feature Engineering
- MODEL 1 - CUSTOMER CHURN PREDICTION
  - Feature Selection And Dimensionality Reduction Using PCA
  - Baseline Model Building
  - Cross Validation And Model Selection
  - Model Evaluation
- MODEL 2 - IDENTIFYING STRONG PREDICTORS OF CHURN
  - Feature selection
  - Handling class imbalance
  - Model building
  - Model evaluation
  - Strategy recommendation to manage customer churn

RESULTS AND DISCUSSION

	model_name	acc_score	roc_score	precision_score	recall_score
1	KNeighborsClassifier	90.27	90.30	83.77	99.80
5	ExtraTreesClassifier	96.85	96.86	96.45	97.26
3	RandomForestClassifier	95.65	95.65	95.18	96.14
2	DecisionTreeClassifier	88.15	88.15	86.37	90.49
0	LogisticRegression	84.59	84.58	85.80	82.77
4	AdaBoostClassifier	83.69	83.68	84.95	81.74



CONCLUSION AND FUTURE WORK

- The obtained results show that our proposed churn model performed better by using machine learning techniques. K-Nearest neighbor produced better accuracy. We identified the main churn factors from the dataset and performed cluster profiling according to their risk of churning. Finally, we provided guidelines on customer retention for decision-makers of the telecom companies.
- The study can be further extended to explore the changing behavior patterns of churn customers by applying Artificial Intelligence techniques for predictions and trend analysis.

REFERENCES

- Gaur Abhishek, Dubey Ratnesh (2018) - Predicting Customer Churn Prediction In Telecom Sector Using Various Machine Learning Techniques - IEEE 2018 International Conference on Advanced Computation and Telecommunication (ICACAT) - [BASE PAPER]
- Ullah Irfan, Raza Basit, Malik Ahmad Kamran, Imran Muhammad, Islam Saif ul, Kim Sung Won (2019) - A Churn Prediction Model using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector - IEEE Access
- Xia Guo-en, Wang Hui, Jiang Yilin (2016) - Application of customer churn prediction based on weighted selective ensembles - IEEE 2016 3rd International Conference on Systems and Informatics (ICSAI)
- Saghir Mehpara, Bibi Zeenat, Bashir Saba, Khan Farhan Hassan (2019) - Churn Prediction using Neural Network based Individual and Ensemble Models - IEEE 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST - 2019)
- Yabas Utku, Cankaya Hakki Candan, Ince Turker (2012) - Customer Churn Prediction for Telecom Services - IEEE 2012 IEEE 36th Annual Computer Software and Applications Conference - COMPSAC 2012
- Zhao Jing, Dang Xing-Hua (2008) - Bank Customer Churn Prediction Based on Support Vector Machine: Taking a Commercial Bank's VIP Customer Churn as the Example - IEEE 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)