Customer Churn Analysis : A look into the data

Saad Hasan[1], Chanpreet Singh[2], Olga Baysal[3], Elio Velazquez[3],

[1]Dept. of Systems and Computer Engineering , [2]School of IT, [3]School of Computer Science

Carleton University

Introduction

What is Customer Churn?
Customer ceases his relationship with the company
Why it is important in Telecom Industry?
In Canada there are lot of telecommunication companies like Bell, Fido, Rogers and others are running to get the big share of the market because of competition. Even losing a single customer a big loss for the company.
Challenges:
Accurate prediction of Customers Churn
Finding the Categories involving customers to churn

Objectives

This project discusses about the churn analysis of telecom company customers and relationship between factors contributing the customer to leave the company as these factors contributing the big loss in the business.
In this project, following are the primary objectives
➢ Finding the factors contributing the customers to Churn
➢ Based on the categories, comparison of different algorithms
➢ Evaluation of existed methods and new methods based on the confusion matrix

Methodology
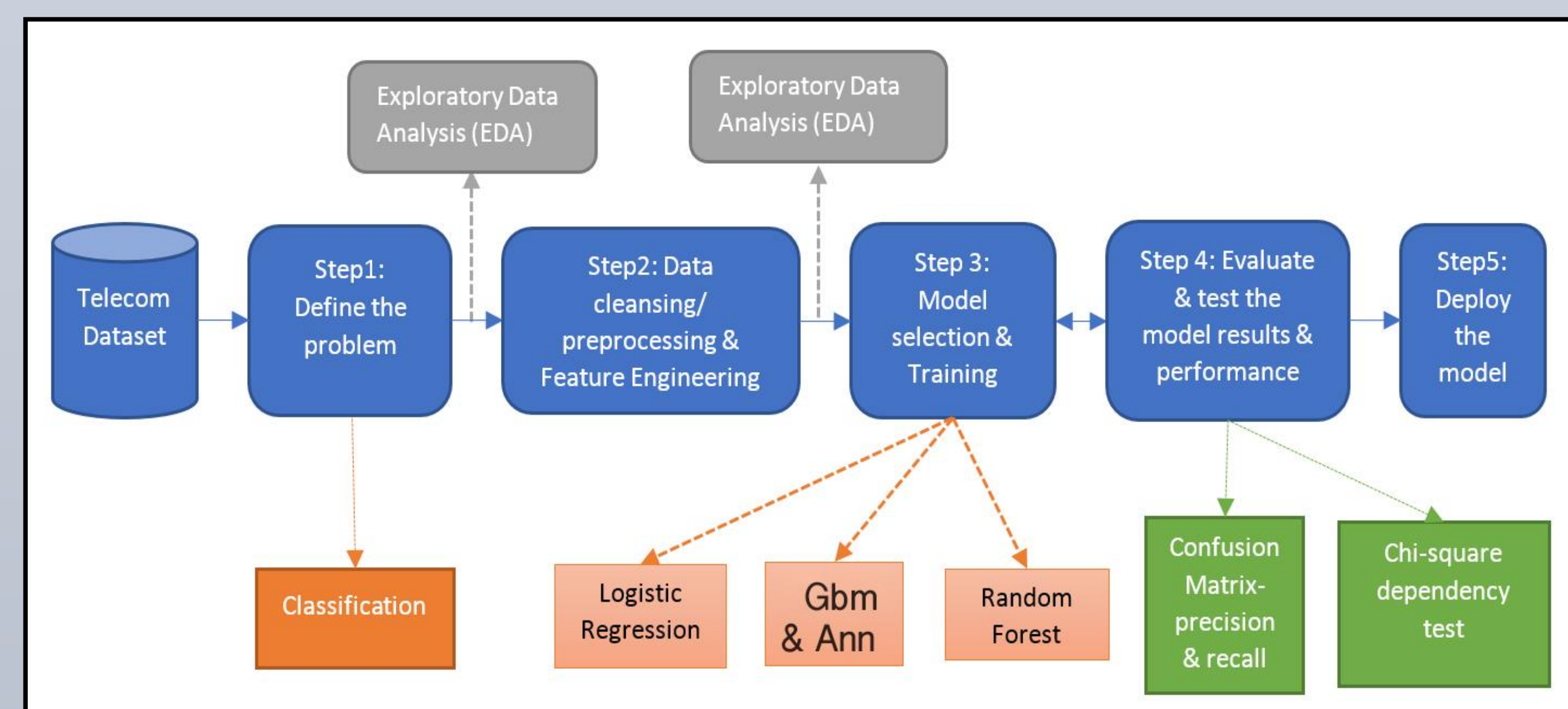
Techniques

Existed methods include Logistic regression and Random forest.

We used Artificial Neural Network and Gradient boosting Machines

$$f(x) = b + xi\ wi$$

Datasets and Limitations

The dataset contains 7022 rows and 21 columns. Features include total Charges, monthly charges, tenure, services availed by customers like Internet Services, Phone Services, type of connections, contract etc.

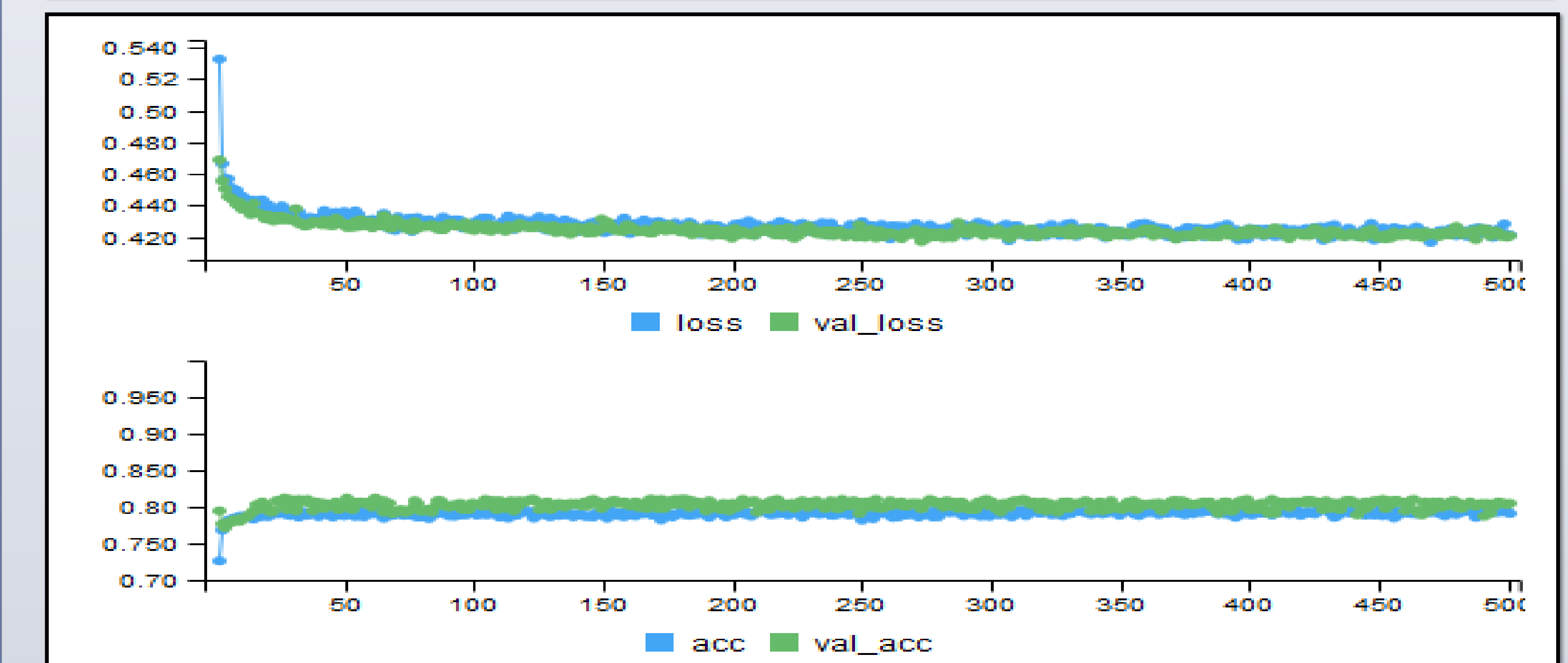The dependent variable is Churn, whether the customer leaves the company or not

The dataset was collected from IBM website published on April 2018
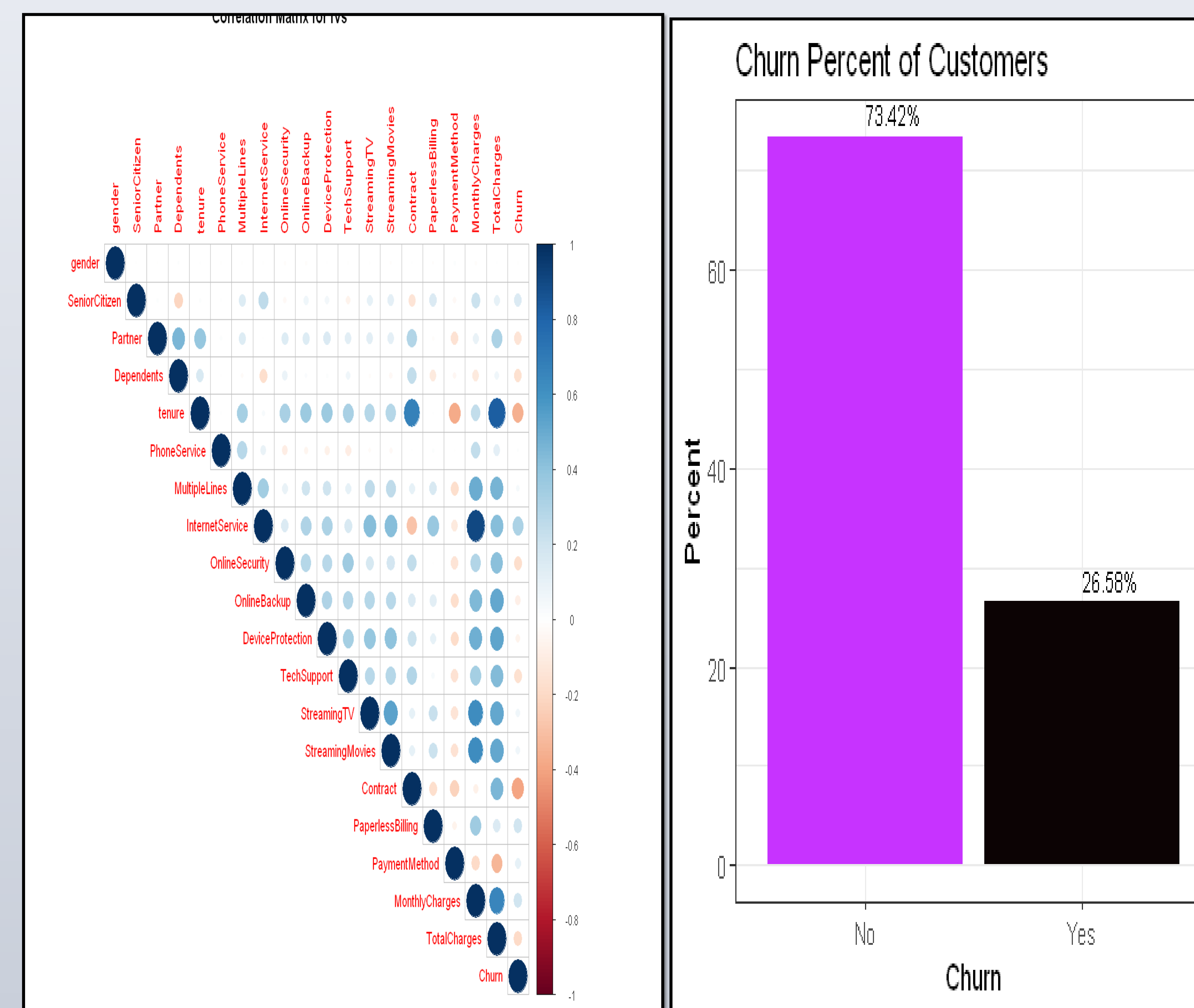
Result
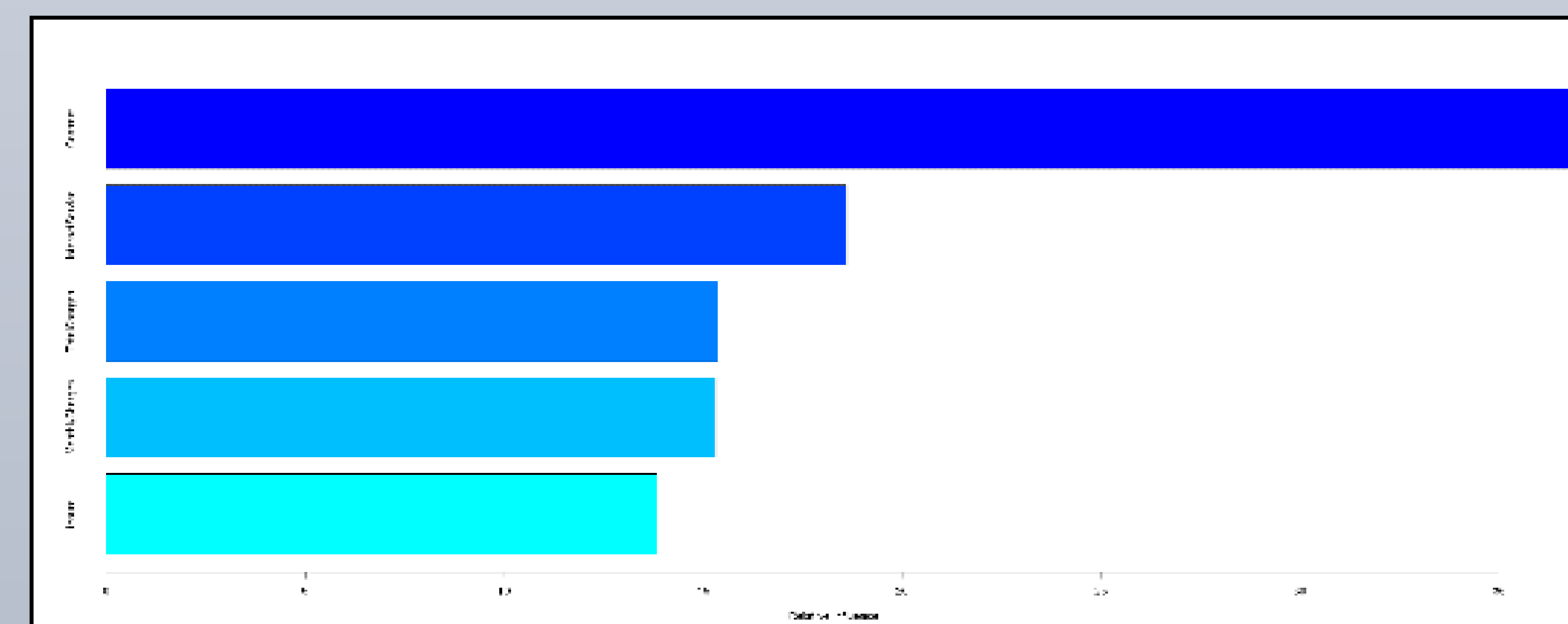
	var	rel.inf
Contract	Contract	36.93163895
InternetService	InternetService	18.60420304
TotalCharges	TotalCharges	15.35249884
MonthlyCharges	MonthlyCharges	15.30416170
tenure	tenure	13.80749747

	Logistic Regression	Gradient Boost Machine	Neural Network	Random Forest
Accuracy	0.77	0.80	0.79	0.78
Precision	0.62	0.65	0.40	0.72
Recall	0.44	0.50	0.70	0.33
Specificity	0.90	0.90	0.80	0.95

Conclusion

Not enough future to predict Churn with good accuracy and recall. In this type of project our focus is to get better recall because every customer who is churn must be identified. Neural network is better in recall as compared to other existed methods used before.

Acknowledgements

We would like to thank our supervisors Olga Baysal and Elio Velazquez for their valuable guidance and motivation which was essential for the progress of our project.

References

[1] Manpreet Kaur and Dr. Prerna Mahajan, "Churn Prediction in Telecom Industry Using R," International Journal of Engineering and Technical Research (IJETR), vol. 3, Issue.5, May. 2015
[2] IBM data analysis on customer support retrieved from https://www.ibm.com/communities/analytics/watson-analytics-blog/sample-datasets