Customer Churn Prediction

End to End Machine Learning Project

Aniket Dubey, IT - 34, Reg. no - 20010816

Purpose and Uses

The purpose of this project is to build a predictive model that can **identify which customers are likely to churn and help the Telco company take proactive measures to retain those customers**. This project can be used by Telco companies or other businesses that have a high customer churn rate to improve customer retention, reduce customer acquisition costs, and increase revenue.

The project can be used in various ways, such as:

- To identify customers who are at a high risk of churn and target them with personalized offers, incentives, or discounts to retain them.
- To improve customer satisfaction by addressing the issues that cause customers to churn and proactively addressing those issues before they become a problem.
- To optimize marketing campaigns by targeting specific customer segments that are more likely to stay with the company.
- To reduce costs by focusing on customer retention rather than customer acquisition, which can be more expensive.

Dataset Availability

The dataset used in the above project is "**WA_Fn-UseC_-Telco-Customer-Churn.csv**". It is publicly available on Kaggle, a platform for data science competitions and projects. The dataset contains information about Telco customers, including demographic information, services used, account information, and whether or not they churned.

The dataset consists of 7,043 rows and 21 columns. It includes both numerical and categorical variables. The target variable is "**Churn**", which is binary and indicates whether the customer churned or not.

Results

Algorithms	Accuracy
Decision Tree Classifier	70.97
Logistic Regression	81.61
Random Forest	79.34
Gradient Boosting Classifier	81.19
Stacked Ensemble Model base models : Decision Tree Classifier, Logistic Regression, Random Forest meta model : Logistic Regression	81.83

Future Scope

Feature Engineering: Although the categorical features were converted to numerical features, there may be more sophisticated ways of encoding categorical data that could improve the model's performance. Additionally, there may be opportunities to create new features from the existing data that could help to better differentiate between customers who are likely to churn and those who are not.

Hyperparameter Tuning: Each of the classification models used in the project has several hyperparameters that can be adjusted to improve performance. By conducting a more systematic search of the hyperparameter space, it may be possible to achieve better results than those obtained using the default parameter values.

Ensemble Learning: While the stacked ensemble model showed some improvement over the individual models, there may be other ensemble techniques that could be explored to further boost performance.

Real-Time Prediction: Once the model is trained and deployed, it could be used to predict churn in real-time. This would allow companies to take proactive steps to retain customers before they decide to leave, which could result in significant cost savings and increased customer loyalty.

Conclusion

The project aimed to predict customer churn in a telecommunications company using machine learning models. The dataset was pre-processed to handle missing values, convert categorical variables into numerical values, and convert the churn variable into binary labels.

Five classification models, including Decision Tree, Logistic Regression, Random Forest, Gradient Boosting, and Stacked Ensemble, were utilized to predict customer churn. The Stacked Ensemble model performed the best with an accuracy of 81.09%, followed by the Gradient Boosting model with an accuracy of 79.09%

Machine learning models can be used to predict customer churn in telecommunications companies with reasonable accuracy. The **Stacked Ensemble model** outperformed the other models and can be used for future predictions. Additionally, the project suggested that incorporating additional data such as customer demographics or usage patterns could further improve the accuracy of the prediction model.

References

1] Huang, B., Kechadi, M.T. and Buckley, B., 2012. Customer churn prediction in telecommunications. *Expert Systems with Applications*, *39*(1), pp.1414-1425.

2]Lazarov, V. and Capota, M., 2007. Churn prediction. *Bus. Anal. Course. TUM Comput. Sci*, *33*, p.34.

3] Xie, Y., Li, X., Ngai, E.W.T. and Ying, W., 2009. Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*, *36*(3), pp.5445-5449.

4] Tsai, C.F. and Lu, Y.H., 2009. Customer churn prediction by hybrid neural networks. *Expert Systems with Applications*, *36*(10), pp.12547-12553.

5] Wei, C.P. and Chiu, I.T., 2002. Turning telecommunications call details to churn prediction: a data mining approach. *Expert systems with applications*, *23*(2), pp.103-112.