Report on Data Analysis and Modeling

Rambabu Alakuntla

April 17, 2024

1 Summary of Exploration

The data analysis revealed a strong correlation between user engagement metrics and the likelihood of user adoption. Users who had at least three visits within a seven-day period showed a significantly higher adoption rate. Additionally, certain features such as creation source and email domains were found to impact user adoption.

2 Preprocessing Steps

1. Identified and merged adopted user information based on engagement data.
2. Calculated the number of days since user creation to understand user behavior over time.
3. Dropped unnecessary columns and performed one-hot encoding on categorical features to prepare the data for modeling.

3 Feature Engineering Steps

1. Created new features to capture the interaction between user engagement and creation source.
2. Engineered features to quantify the frequency of user visits within specific time intervals.

4 Model Selection

Utilized a Random Forest Regression model with 100 estimators to predict user adoption. Random Forest was chosen for its ability to handle complex interactions between features and provide feature importance rankings.

5 Conclusion

The Random Forest Regression model successfully identified key features influencing user adoption, such as creation source and frequency of user visits. The model can be used to predict user adoption rates and guide targeted marketing strategies to improve user engagement.

6 Reference Link

https://colab.research.google.com/drive/1V8qwysnN4LIRKzLTPMKbvpDcwe-0jK3z?usp=sharing