Tensor-IOT task Report

Problem Statement :- Need to identify variables contributing to converting users into adopted users and perform analysis to describe an analysis that shows non-technical stakeholders what variables and conditions are associated with user adoption.

Summary :- The random forest classifier identified several features contributing to user engagement. User lifespan emerged as the most influential predictor, indicating longer user engagement durations correlate with higher engagement levels. Other notable factors include being invited by another user and the organization to which the user belongs, Additionally, factors such as email domain and creation source play minor roles in user adoption.

Preprocessing & Feature engineering steps :- The steps taken for preprocessing & feature engineering are as follows.

- Basic data exploration to understand the data and to check the data types and missing values in data.
- Performed unix time data and string time data conversion to datetime format and created function to generate a feature user_adopted.
- Used imputation techniques to impute missing value and performed data type conversion as needed.
- Created new features like life span, domain from existing feature to understand the relation.
- Kept necessary columns and performed Label encoding to convert categorical features into numerical features.
- Converted data into train-test split for model training and testing.

Model selection :- Used Random Forest classifier for the ml solution, as we had a small dataset and mix of categorical and numerical data. Also it works well with class imbalance and is easy to implement. We had achieved 97.26% accuracy which defined that our model convincingly able to classify used will be adopted or not.

Conclusion :- Analysis using the random forest classifier showcased the user lifespan as the primary indicator of whether the user will be adopted or not. Invited by user id and organization id has also significant impact while email and creation source has minor impact on user adoption. Also analysis shows that most active users spend more time on the platform than non-adopted users. Around 60% of active users exceed the usage duration of non-adopted users.