HappyDB – Discover the Happy Moments

PRESENTED BY –

HANSRAJ PABBATI

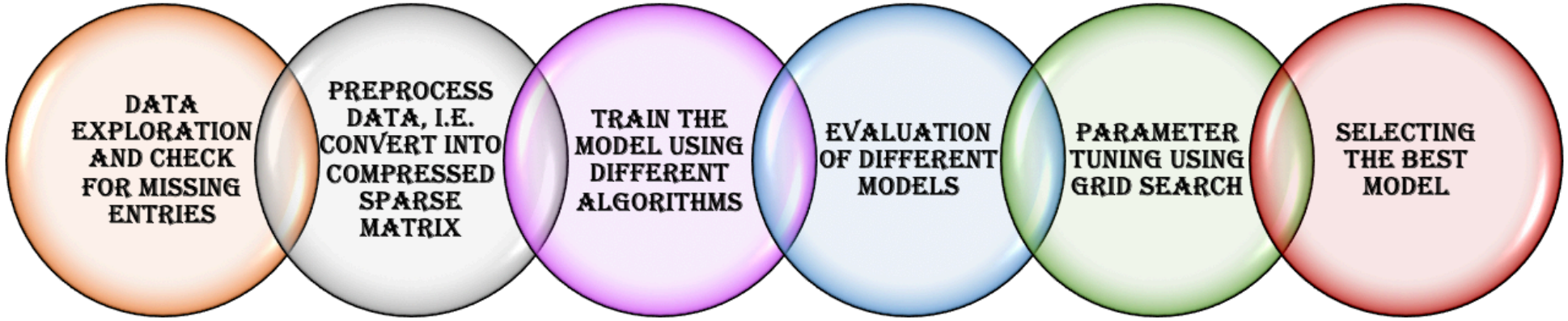
NEHA BINDLE

SHREY PATEL

Objective

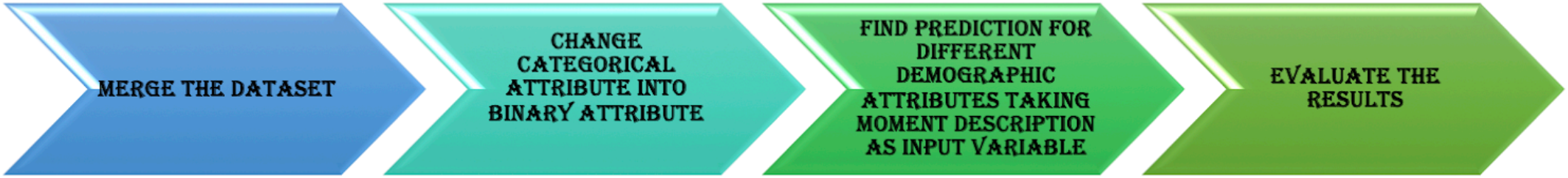
- To predict the category of happiness based on the description of the moment.

- To predict demographic variables such as Age, Country, Marital status, Parenthood, the Reflection period, and Gender depending on the description of a happy moment.



System Design

FOR PREDICTING THE HAPPINESS CATEGORY BASED ON MOMENT DESCRIPTION



System Design

FOR PREDICTING THE DEMOGRAPHIC ATTRIBUTES BASED ON MOMENT DESCRIPTION

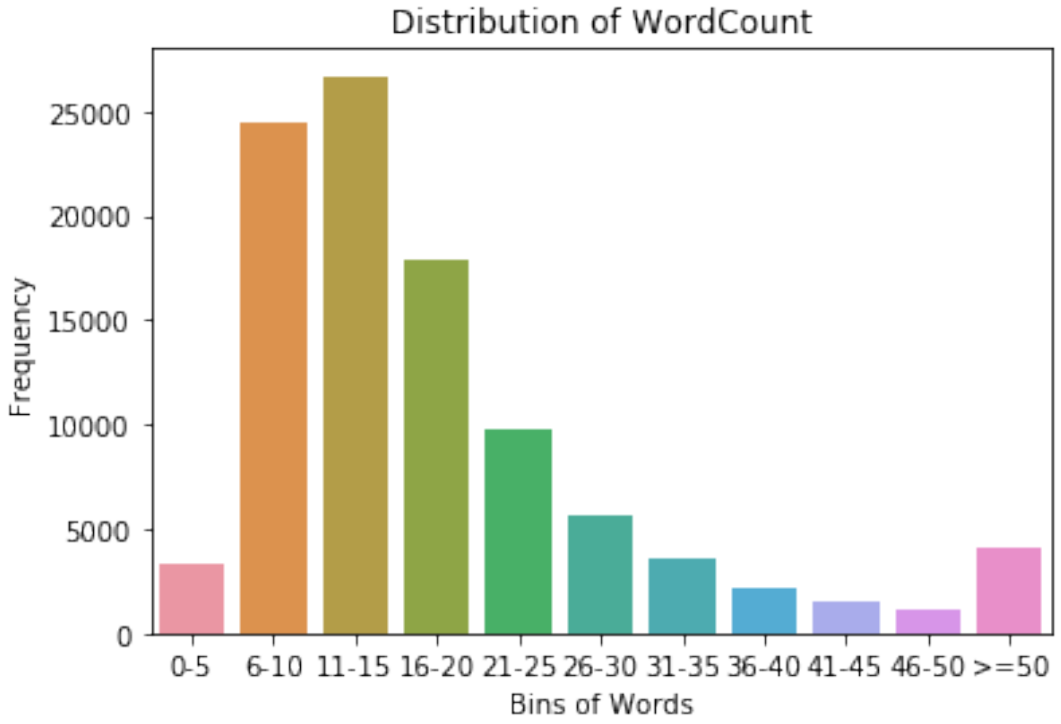
Data Preprocessing

Text Data

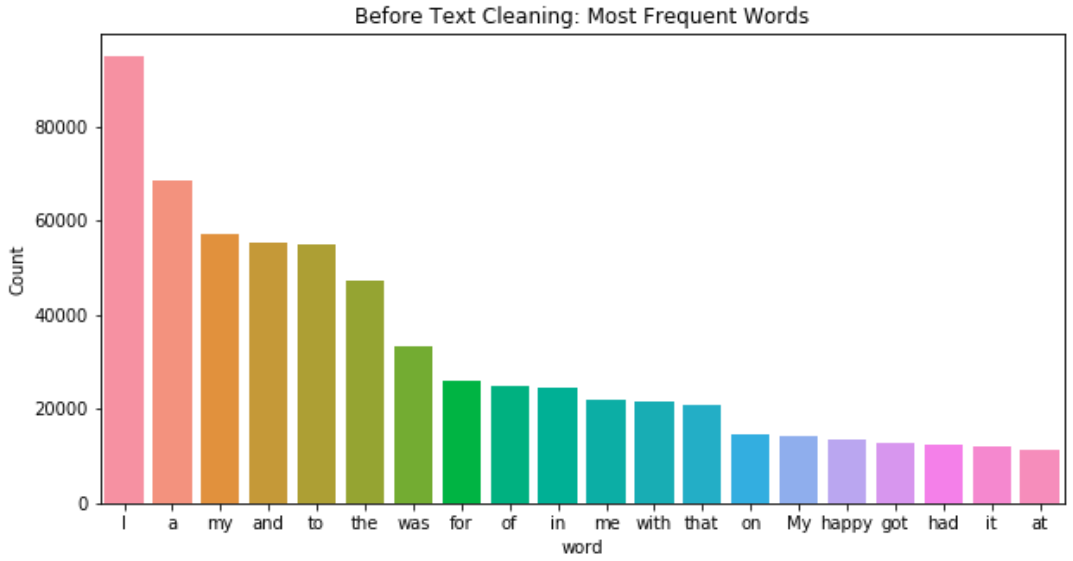
- Removal of Regex, all text into lower case
- Dropping null values
- Removal of Noisy and less meaningful words
- Tokenization, Lemmatization, CSR Matrix

Demographics Data

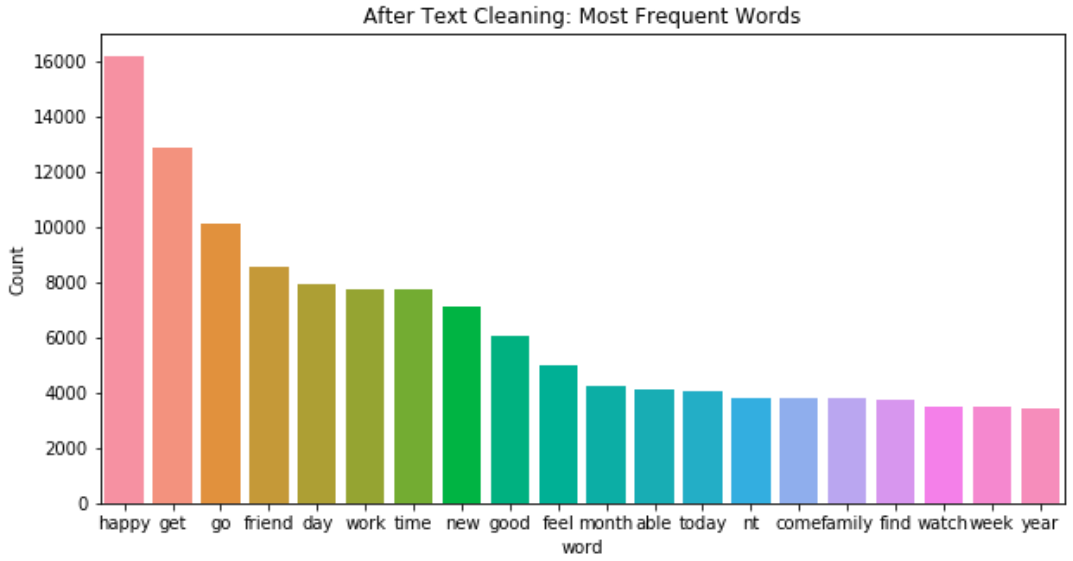
- Age less than 25 as 0, else 1
- Parenthood no as 0 and yes as 1
- Single, divorced, Widower or separated as 0, Married as 1
- Reflection Period 24 hours as 0, and 3 months as 1
- Gender Male as 0 and Female as 1
- Country USA as 1, else 0

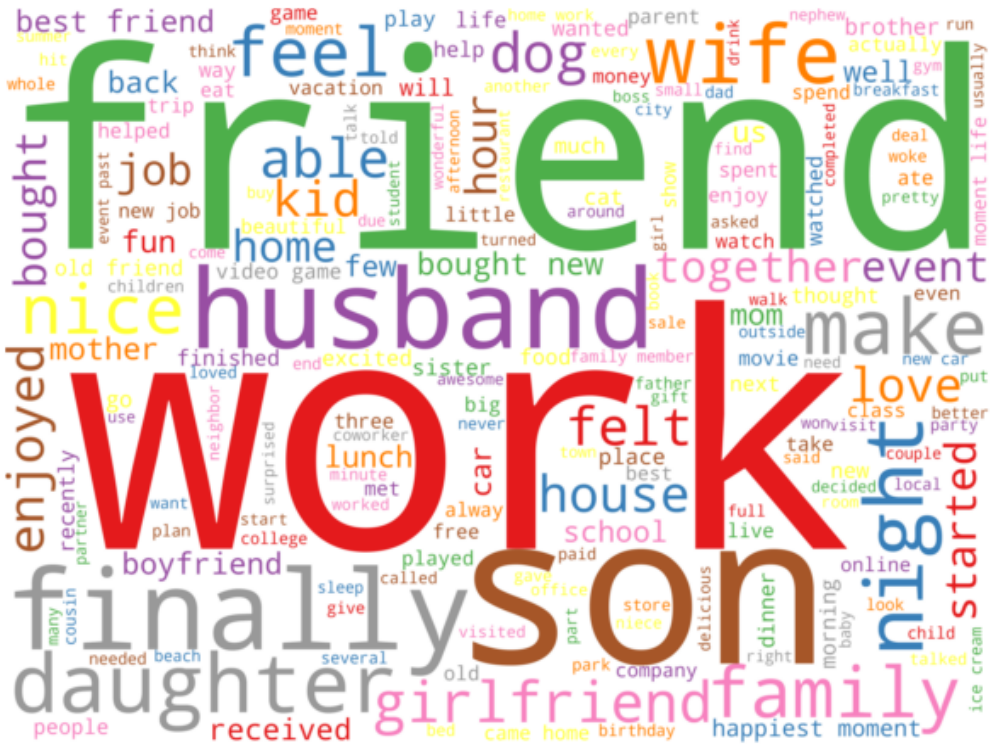


Data Exploration



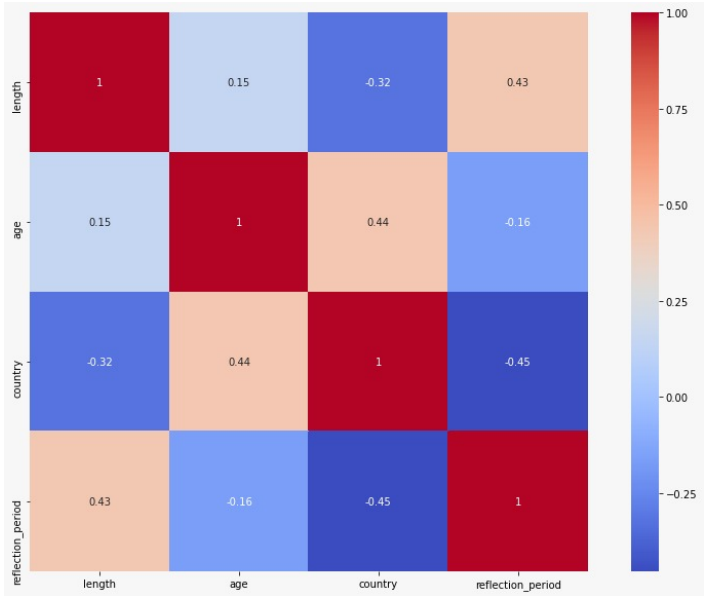
Before Text Cleaning

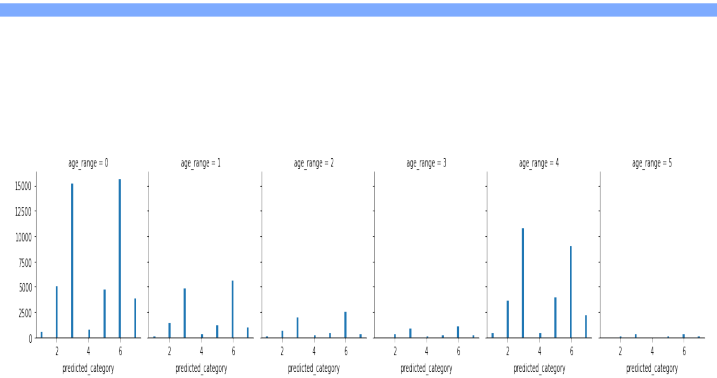


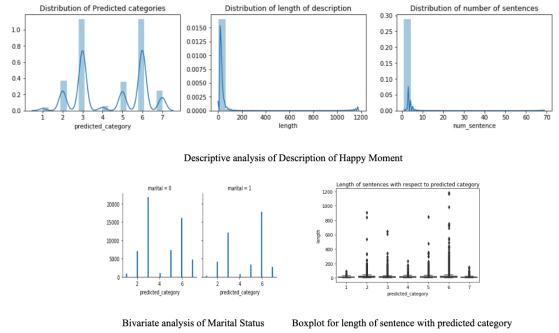


After Text Cleaning

Metadata Analysis







Descriptive analysis of Description of Happy Moment

Bivariate analysis of Marital Status

Boxplot for length of sentence with predicted category

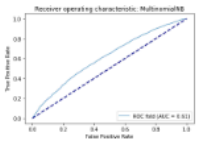
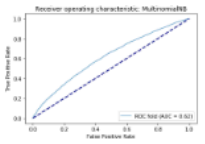
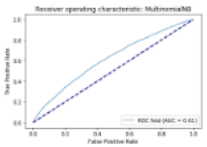
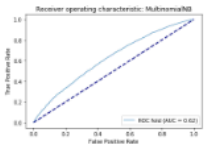
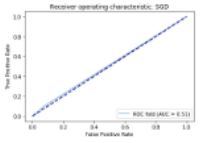
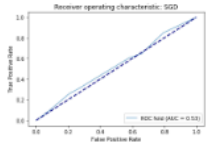
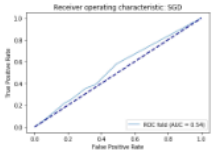
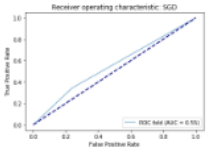
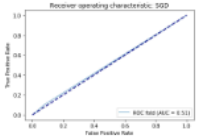
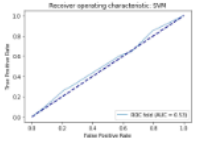
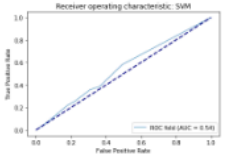
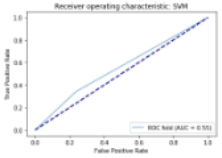
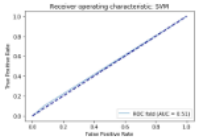
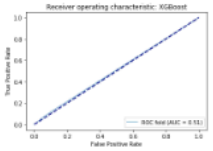
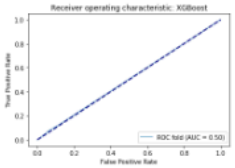
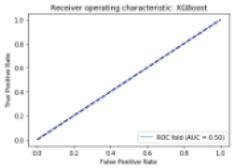
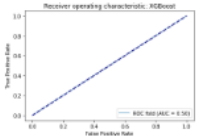
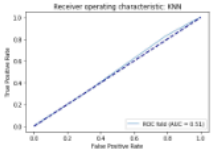
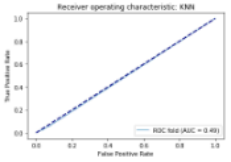
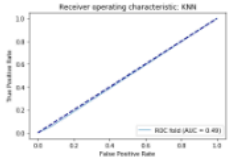
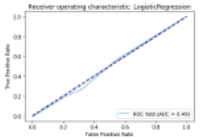
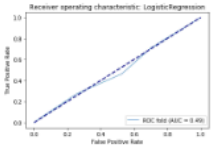
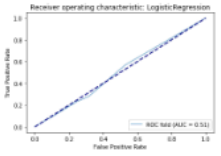
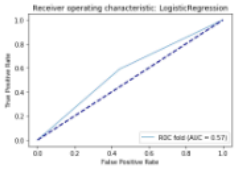
	importance
sentimentPolarity	0.273675
sentimentSubjectivity	0.272600
length	0.269831
predicted_category	0.068483
marital	0.054031
parenthood	0.047000
reflection_period	0.014380

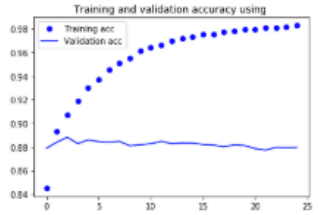
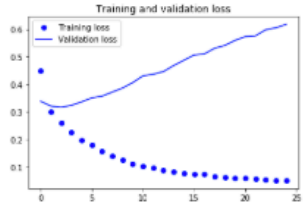
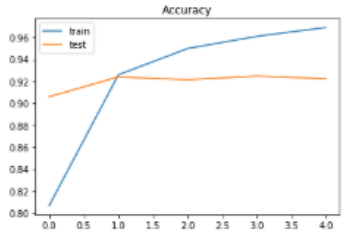
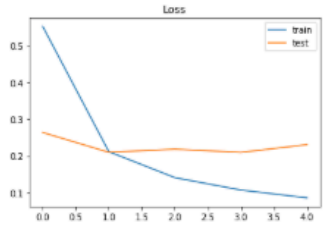
Analysis of Different models

Model Used	F1 score Regular	F1 Score with SMOTE	F1 Score with ROS	F1 score with Binary Conversion
Multinomial Naive Bayes	0.39	0.36	0.36	0.60
SGD	0.35	0.24	0.30	0.52
SVM	0.31	0.22	0.30	0.48
XGBoost	0.34	0.33	0.32	0.43
KNN	0.34	0.13	0.33	0.43
Logistic Regression	0.32	0.30	0.32	0.58

Models for predicting the Happiness Category based on Moment Description

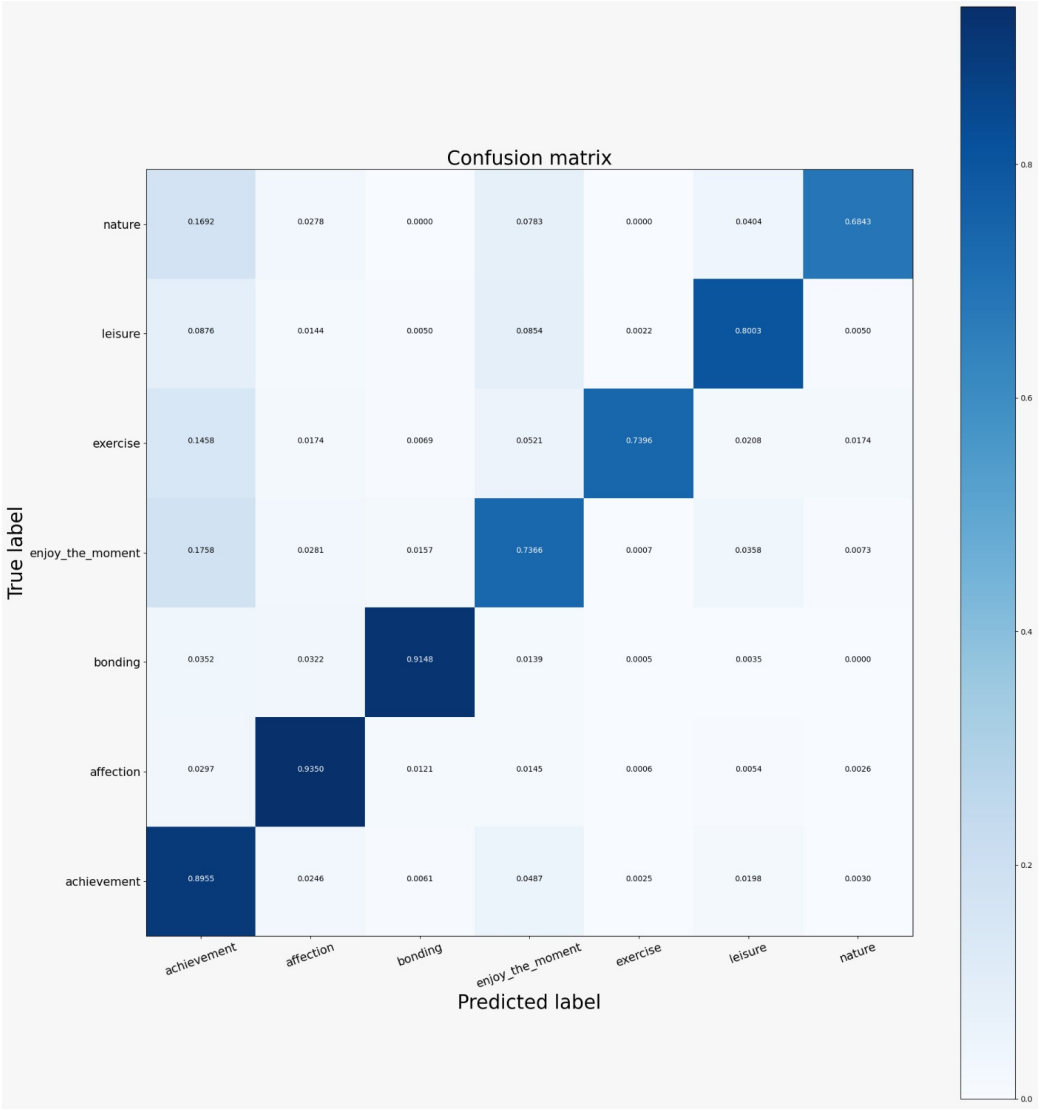
ROC CURVE

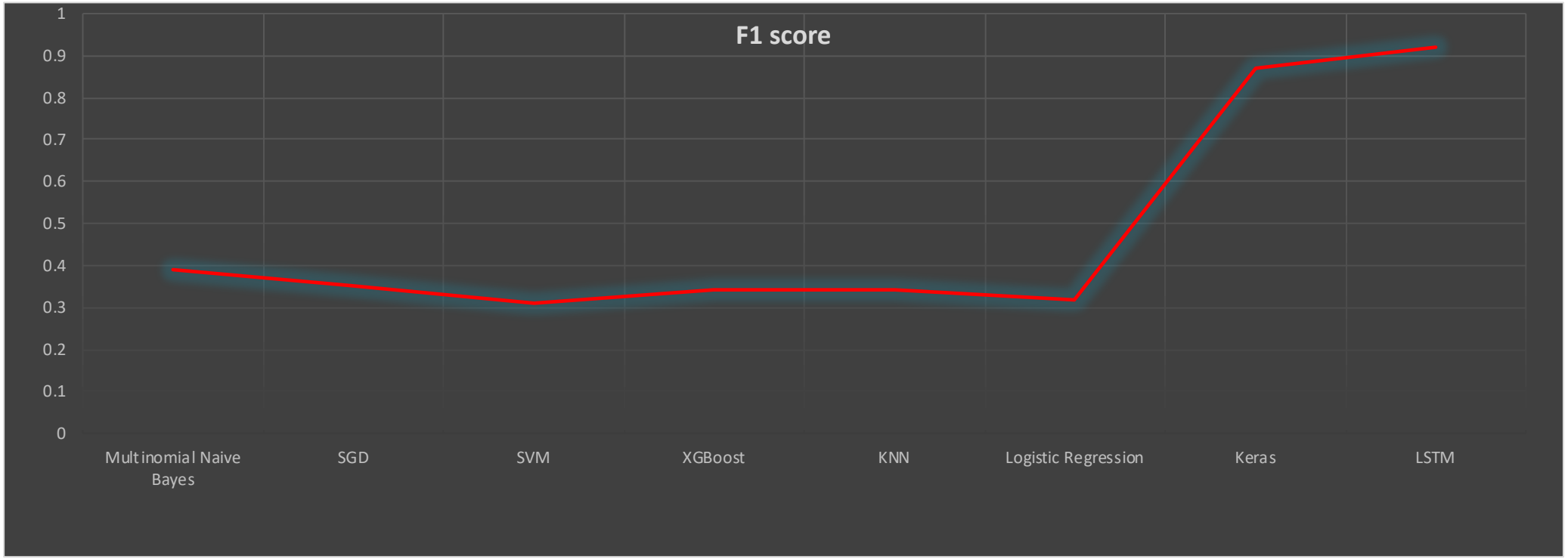
Model Used	ROC Curve Regular	ROC with SMOTE	ROC with ROS	ROC Binary Conversion
Multinomial Naive Bayes				
SGD				
SVM				
XGBoost				
KNN				
Logistic Regression				

Model Used	Test Accuracy	Training and validation Accuracy	Training and validation Loss
Keras	0.87		
LSTM	0.92		

Deep learning models for predicting the Happiness Category based on Moment Description

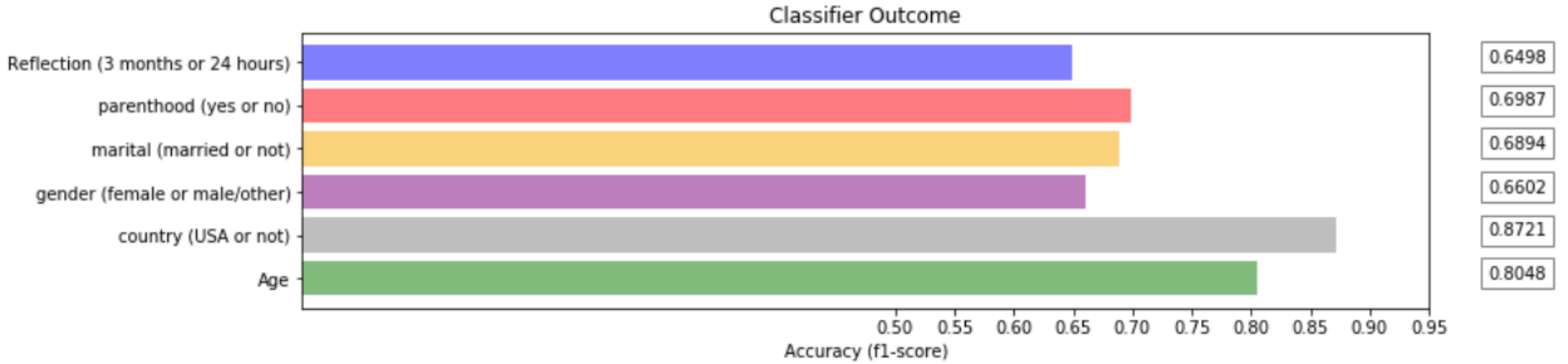
```
I made the most delicious meal for my significant  ...
Actual label:affection
Predicted label: affection

I spent time with colleagues at a work conference  ...
Actual label:bonding
Predicted label: bonding

I donated a bunch of old books I had to the local  ...
Actual label:enjoy_the_moment
Predicted label: enjoy_the_moment

I attended the wedding of my cousin. ...
Actual label:affection
Predicted label: affection

I found $50 in my winter jacket ...
Actual label:achievement
Predicted label: achievement

I Went to the Dollar Store earlier and was able to ...
Actual label:achievement
Predicted label: achievement

I got a big lead at work and was recognized for it ...
Actual label:achievement
Predicted label: achievement

Watch a movie of terror at home, eating cotufas an ...
Actual label:leisure
Predicted label: leisure

I got a really nice desert last night and a fidget ...
Actual label:achievement
Predicted label: achievement

last month i went a tour to banglore, and i enjoye ...
Actual label:enjoy_the_moment
Predicted label: enjoy_the_moment
```




F1 score
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Multinomial Naive Bayes
SGD
SVM
XGBoost
KNN
Logistic Regression
Keras
LSTM

Conclusion

	Age	Country	Gender	Marital	Parenthood	Reflection_period
Accuracy	0.80	0.87	0.66	0.68	0.69	0.64



Multinomial Naive Bayes Model for predicting the Demographic attributes based on Moment Description

Challenges

- Noisy Data
- Interpretation of Categorical attributes
- Poor accuracy in the beginning
- Predicting importance of demographic attributes

Any Questions?