OffensEval

**Team No : 18**

Archit Kumar

Arnav Kapoor

Neeraj Barthawal

Shruti Chandra

Assigned TA : Vivek Anand

## Project URL : <https://drive.google.com/drive/folders/1GalFownguSIn9XJSZDMJZul4oFp_ak7s?usp=sharing>

# Overview

A system to classify a tweet as either offensive or not offensive . It involves following three subtasks which are hierarchical in nature :

## 

## Subtasks

* Sub-task A: Offensive language identification; [Offensive: OFF, Not Offensive: NOT]
* Sub-task B: Automatic categorization of offense types; [Targeted: TIN, Untargeted: UNT]
* Sub-task C: Offense target identification. [Individual: IND, Group: GRP, Other: OTH]

## Dataset used : OLID v1.0 dataset

# Pre-Processing Steps

* Decontraction - (can’t -> can not , won’t -> will not )
* Stop Words Removal
* Case folding
* Tokenization and Lemmatization

# Additional Features Used

* Number of Hashtags
* Number of User Mentions
* Length of Tweet
* Number of URLs and Emojis.

# Representations Formed

* Bag of Words
* TF-IDF
* Average Word2Vec (300 dimensions)

# Models Used

* Logistic Regression
* Naive Bayes
* SVM

# Task Description

* Task A - Binary ( OFF , NOT)
* Task B - Binary (TIN , UNT)
* Task C - Multiclass (GRP , IND, OTH)

Results

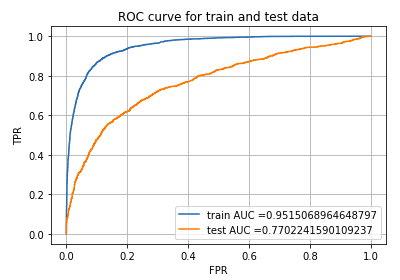
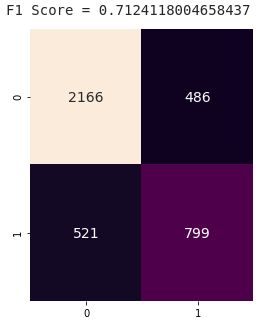
# **TASK A**

F1- Macro Scores

|  |  |  |  |
| --- | --- | --- | --- |
|  | Bag of Words | Tf-Idf | Avg Word2Vec |
| Logistic Regression | 0.712 | 0.710 | 0.649 |
| Naive Bayes | 0.705 | 0.658 | 0.591 |
| SVM | 0.700 | 0.702 | 0.643 |

Best Result - Logistic Regression (Bag of Words)

Confusion Matrix and ROC curve



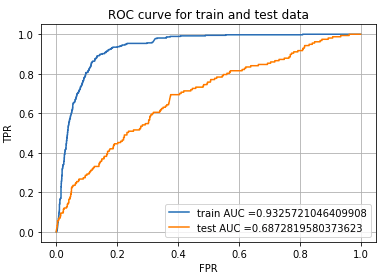
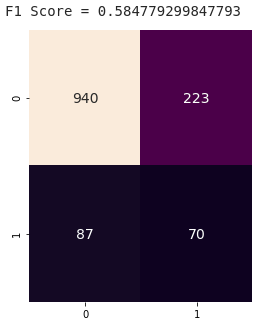
# **TASK B**

F1- Macro Scores

|  |  |  |  |
| --- | --- | --- | --- |
|  | Bag of Words | Tf-Idf | Avg Word2Vec |
| Logistic Regression | 0.560 | 0.584 | 0.555 |
| Naive Bayes | 0.530 | 0.5489804 |  |
| SVM | 0.56091 | 0.55884 | 0.538198 |

Best Result - Logistic Regression (TF-IDF)

Confusion Matrix and ROC curve



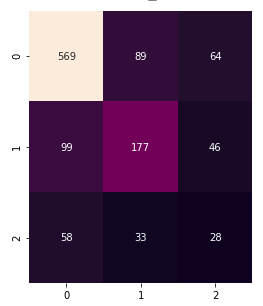
# **TASK C**

Task C is a multiclass classification problem . Hence following score are based on ‘micro’ averaging of F1 scores of each individual class.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Bag of Words | Tf-Idf | Avg Word2Vec |
| Logistic Regression | 0.70991522 | 0.72871 | 0.6000737 |
| Naive Bayes | 0.6603611 | 0.6595012 | 0.62080825 |
| SVM | Linear: 0.60705  Rbf: 0.67411 | Linear: 0.63628  Rbf: 0.59587 | Linear : 0.5932  Rbf: 0.5666 |

Best Result - Logistic Regression (TF-IDF)

Confusion Matrix



## 

* Used class weights to handle class imbalance issues .
* Regularization : L2
* In logistic regression : used one vs rest method for multiclass classification .
* In SVM :

1. Linear SVC : linear kernel with one vs rest based classification
2. SVC : ‘rbf’ kernel with one vs one approach

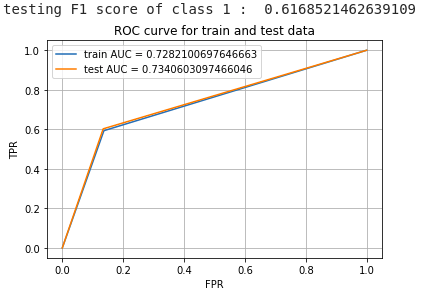
## Result for individual classes :

No model perform well on individual classes , due to small amount of data and class imbalancing . We will try to improve this classification section in next deliverable .

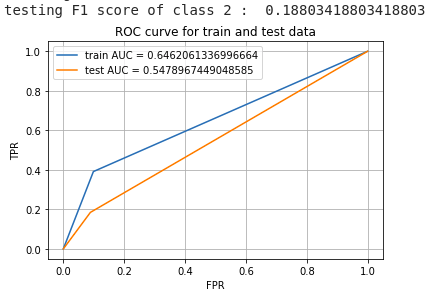
**Class 0 : IND ( Individual )**

## 

**Class 1 : GRP ( group )**



**Class 2 : OTH ( other )**



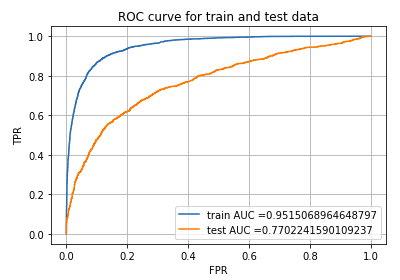
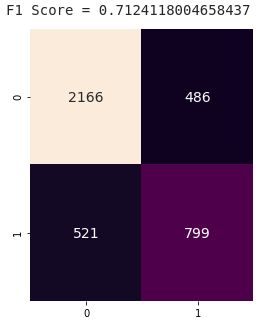
# Additional Features Analysis

For Task A we analysed the results with and without the additional features (# of hashtags , # of mentions , # of URls etc).

## With Additional Features

Best Result - Logistic Regression (Bag of Words)

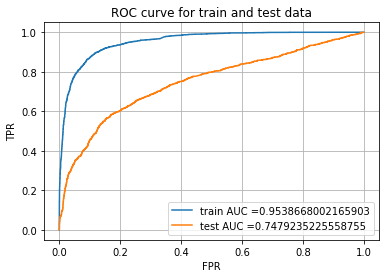
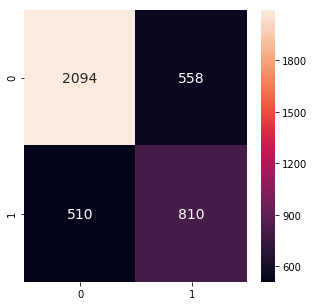
Confusion Matrix and ROC curve



## Without Additional Features

Logistic Regression (Bag of Words)

F1 Score = 0.6997411121983039



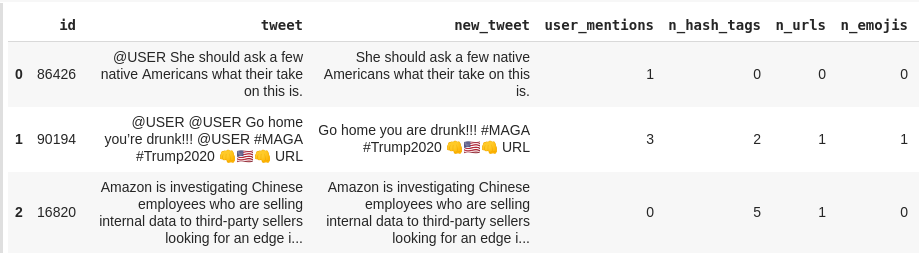
**Thus we see that the additional features do help as without them F1 Score drops marginally from 0.712 to 0.699.**

User Query

We have also added user query (new tweets) as input , the same can be seen in Naive\_Bayes Task B (Targeted vs Untargeted). We will replicate this across all the tasks.

# 

# Tweets after Preprocessing :



# Misclassification examples :

## Task A :

Red: original not offensive (0), classified as offensive (1)

Green: original offensive (1), classified as not offensive (0)

1. @USER @USER But the liberals do
2. @USER @USER Sshh 🙊 she is lying😏😅😭😂
3. @USER Goodell is the worst commissioner of any sport from any time in history
4. @USER I'm frothing over all of it so far.. 🤤🤤 ..the goodest shit. 😏

## Task B :

Red: original Targeted insult (0), classified as untargeted (1)

Green: original untargeted (1), classified as Targeted insult (0)

1. @USER @USER If this yutz was any dumber she'd have to be watered twice a week!
2. @USER His mouth goes to one side all smirky and shit and I don’t like it.
3. @USER @USER Son please don’t make me cry today because I never saw this shit
4. @USER He tihnks he is telling the big lie.

## 

## Task C :

Red: original individual insult (0), classified as group (1)

Green: original individual insult (0), classified as other (2)

Purple: original group insult (1), classified as individual (0)

Cyan: original group insult (1), classified as other (2)

Light Blue: original other insult (2), classified as individual (0)

Orange: original other insult (2), classified as group (1)

1. @USER Yeah. That's kinda the fashy thing right now. The Proud Boys scumbags wear tshirts saying Pinochet did nothing wrong" And Antifa are the ones who are violent? Please. It's been an actual fascist every time this has been brought up in my experience. Like dude here, who stopped.."

2. @USER @USER I think the pope and some others should be prosecuted for covering up and protecting child rapists!

3. @USER @USER Beware of rightist false flags like the narrative of the OK sign being a white nationalist dog whistle. It's not. #TheResistance #Resist #BashTheFash #Antifa #SmashFascism URL

4. @USER Its clear you wouldn't listen anyways since you are more interested in arguing over the non existent racism of that tweet rather than calling out the ignorant white girl on display

1. @USER @USER Mine go drafts too. @USER stop this crap I'm gathering proof your censoring conservative .
2. @USER Zimmerman belongs in prison for killing Travon Martin. Florida needs to change its gun control laws.

Misclassified examples shows that some samples are misclassified because they were labeled wrong .

# Any change in methodology from scope document :

We adhered to the original plan and implemented the models as planned earlier. The plan for the first deliverable was to have the pre-processing done , have features in place to try out the basic ML models.

# Final Deliverables :

For final deliverable we will include deep learning models ( CNN , LSTM ) for each task using different representation of tweets data .