<u>IS6713 Project Report</u>

<u>Team Members:</u>

Gagana Uday Kumar - WOV796
Niharika Bandlamudi - HBZ194
Nupoor Karnik - JHR497
Munivenkataparthasai Madallapalli -PUA528

<u>Selection of data, topic and guidelines:</u>

- **The data was City of San Antonio’s reddit dataset.**
- **The focus of our study was if ‘People of San Antonio used technology and specifically ‘Ride Share Apps (Via, Uber, Lyft, Ridesharing etc.)’, ‘Food Deliveries (Uber Eats/Deliveries, Doordash, GrubHub, HEB, Whole Foods, etc.)’, ‘Online Shopping (Amazon, Walmart, FB marketplace, etc)’’ as we were interested in knowing these particular things.**

<u>Process of Annotations:</u>

Part 1- 25 annotations given to 2 annotators. Any disagreement was discussed and corrected.
Part 2- later 975 comments were sent to annotators to get them annotated. Any disagreement was discussed and corrected.

<u>Process and results of Modeling:</u>

After the annotation process was complete, we appended the data to prepare the list then we used kappa statistics to get the agreement score.

Following are the scores of kappa agreement:
Cohen's Kappa Agreement Score Technology 0.5268
Cohen's Kappa Agreement Score Ride Share 0.3986
Cohen's Kappa Agreement Score Food Delivery 0.3981
Cohen's Kappa Agreement Score Online Shopping 0.4597

- According to Cohen's kappa agreement score we had performed model selection for training and prediction.
- After reading the raw data into *Final_Annotationsby_Detectives_GSD.csv* we converted it into two lists for training the data i.e., X_txt will be a list of strings with features. y will list 0's and 1's with labels or classification. Later we split the data into train and test sets.
- Firstly, for base modeling we used a linear model of **LogisticRegression** with feature extraction of **CountVectorizer** for modeling with ngram_range (1,1).
- We converted the data into matrices and later converted it into an array.

- Later we initialized gridsearchCV with f1_macro and found the validation score then we calculated the precision recall and f1_score with average as “macro and micro parameters”. **Following are the scores:**

    - Macro Validation Score F1: 0.4110
    - Macro Test Score F1: 0.4007
    - Macro_Score Precision: 0.4137
    - Macro_Score Recall: 0.3926
    - Gold Standards- Technology Macro_Score F1: 0.8201
    - Gold Standards- Ride Share Macro_Score F1: 0.2069
    - Gold Standards- Food Delivery Macro_Score F1: 0.3492
    - Gold Standards- Online Shopping Macro_Score F1: 0.2264
    - Micro Validation Score F1: 0.5229
    - Micro Test Score F1: 0.5010
    - Micro_Score Precision: 0.5708
    - Micro_Score Recall: 0.4464
    - Gold Standards- Technology Micro_Score F1: 0.8319
    - Gold Standards- Ride Share Micro_Score F1: 0.2078
    - Gold Standards- Food Delivery Micro_Score F1: 0.2778
    - Gold Standards- Online Shopping Micro_Score F1: 0.1522

- Looking at the above results it is said that the model performance is good, macro validation score is 41% and micro validation score is 52%. Also, the model predicts the best for technology with approximately 83%.

- Secondly, we also tried base modeling using a linear model of **LogisticRegression** with feature extraction of **TfidfVectorizer** for modeling with ngram_range (1,5).
- We converted the data into matrices and later converted into an array and added the parameters.
- Later we initialized gridsearchCV with f1_macro and found the validation score then we calculated the precision recall and f1_score with average as “macro and micro parameters”. **Following are the scores:**

    - Macro Validation Score F1: 0.2106
    - Macro Test Score F1: 0.2612
    - Macro_Score Precision: 0.4455
    - Macro_Score Recall: 0.2639
    - Gold Standards- Technology Macro_Score F1: 0.7939
    - Gold Standards- Ride Share Macro_Score F1: 0.0377
    - Gold Standards- Food Delivery Macro_Score F1: 0.0988
    - Gold Standards- Online Shopping Macro_Score F1: 0.1143
    - Micro Validation Score F1: 0.4964
    - Micro Test Score F1: 0.4881
    - Micro_Score Precision: 0.6494
    - Micro_Score Recall: 0.3910
    - Gold Standards- Technology Micro_Score F1: 0.7839

- Gold Standards- Ride Share Micro_Score F1: 0.0377
- Gold Standards- Food Delivery Micro_Score F1: 0.0533
- Gold Standards- Online Shopping Micro_Score F1: 0.0968

- Looking at the scores the model doesn't predict well compared to the previous model as the macro validation score is 21% and macro validation score is 49% and the technology score is approximately 79%.
- With the help of **lexicon classifiers,** we declare each class i.e Exclamation point, Tech word Count, Online App (includes ride-share and online shopping) and food delivery and their count.
- We used a linear model of LogisticRegression with feature extraction of CountVectorizer for modeling with count of exclamation marks with ngram_range (1,1) and lexicon features labels for hstack.
- We converted the data into matrices and later converted it into an array.
- Later we initialized gridsearchCV with f1_macro and found the validation score then we calculated the precision recall and f1_score with average as "macro and micro parameters".
- **Following are the scores for all the classes:**

FEATURE	**Exclamation Marks**	**Tech**	**Online**	**Food**
Macro Validation Score F1	0.4115	0.3796	0.2504	0.3429
Macro Test Score F1	0.399	0.3429	0.2398	0.3216
Macro_Score Precision	0.4113	0.3956	0.4013	0.4368
Macro_Score Recall	0.3926	0.324	0.2634	0.2927
Gold Standards- Technology Macro_Score F1	0.8099	0.823	0.7535	
Gold Standards- Ride Share Macro_Score F1	0.2069	0.1176	0.0741	
Gold Standards- Food Delivery Macro_Score F1	0.3548	0.2752	0.046	0.1429
Gold Standards- Online Shopping Macro_Score F1	0.2243	0.1556	0.0857	
Micro Validation Score F1	0.5208	0.5212	0.04967	0.5131
Micro Test Score F1	0.4982	0.501	0.4807	0.482
Micro_Score Precision	0.5292	0.6089	0.5842	0.6196

Micro_Score Recall	0.4706	0.4256	0.4083	0.3945
Gold Standards- Technology Micro_Score F1	0.8151	0.8197	0.7331	
Gold Standards- Ride Share Micro_Score F1	0.1882	0.1818	0.0385	
Gold Standards- Food Delivery Micro_Score F1	0.3471	0.2353	0.0286	0.1443
Gold Standards- Online Shopping Micro_Score F1	0.1961	0.1266	0.069	

- From the above mentioned scores, it can be seen that maximum use of technology by people of SA is done for food delivery apps and then for online shopping and ride share.

Finally, we did Manual analysis:

For example:

Tweet: If you wish to stick with iPhone, the iPhone 6S is still currently supported with iOS 15. You can get one used for about $75 on eBay and maybe even cheaper on a local marketplace app
['Gold Standards- Technology', 'Gold Standards- Ride Share', 'Gold Standards- Food Delivery', 'Gold Standards- Online Shopping']
Ground-Truth Class: [1, 0, 0, 1]
Lexicon Exclamation Model Prediction(micro f1): [0 1 1 0]
Lexicon Exclamation Model Prediction(macro f1): [0 1 1 0]
Lexicon Technology Word Count Model Prediction(macro f1): [0 0 1 0]
Lexicon Technology Word Count Model Prediction(micro f1): [1 0 1 0]
Lexicon Online App Word Count Model Prediction(micro f1): [1 0 0 0]
Lexicon Online App Word Count Model Prediction(macro f1): [1 0 0 0]
Lexicon Food Delivery App Word Count Model Prediction(micro f1): [1 0 1 0]
Lexicon Food Delivery App Word Count Model Prediction(macro f1): [0 1 1 0]

According to manual analysis it is observed that the model predicts decently well.
Therefore, according to the scores given above we can conclude that the model is a good fit. This indicates that people of SA are users of technology and according to the CountVectorizer feature model we can say that they use it in their daily lives majorly for food delivery followed by rideshare and online shopping. Hence, we can say that food delivery tops the list.