Detection Of Bot Accounts On Social Networks Using Big Data Mining Tools

U17CO002 Kaneesha Gandhi
U17CO023 Harshit Sodagar
U17CO028 Devanshi Bhatia
U17CO074 Jay Rathod

Guided by:
Dr. Dipti P. Rana

Introduction

- ONLINE SOCIAL NETWORKS ARE UBIQUITOUS.
- ATTRACTIVE TARGETS FOR MALICIOUS ENTITIES.
- SOCIAL BOTS ARE CAPABLE OF DOING THINGS WE CANNOT IMAGINE.
- ALL THIS TELLS HOW IMPORTANT IT IS TO IDENTIFY THESE FAKE ACCOUNTS.

Objectives

- TO COLLECT AND PROCESS DATA FROM ONE OR MORE SOCIAL MEDIA PLATFORMS.
- TO IDENTIFY WHETHER A GIVEN ACCOUNT IS A BOT OR A LEGITIMATE USER.

Literature Review

Authors like Bilge[1] worked on preventing automated attacks such as profile cloning and cross-site profile cloning. These were one of the most common identity thefts.

Furthermore, authors like Jin[2], Kontaxis[3] and Mohammed Razaei[9] analyzed the extensive social network patterns of any given user and compared the similarities of an input identity to the user over various networks.

Literature Review

Claudia[6] and Ragusa[6] put forth a technique which utilized sampling of non-uniform features inside a machine learning algorithm by the adaptation of random forest algorithm to recognize spammer insiders.

Lastly, authors like Deniz[7] and Gharge, collected Twitter datasets and used them for the detection of spam accounts. Deniz utilized the machine learning approach of Naive Bayes learning algorithm, before and after discretization of data. Gharge initiated a method, which was classified on the basis of two new features.

Proposed Framework

INTRODUCTION

- GOAL IS BINARY CLASSIFICATION : BOT OR LEGITIMATE.

- ALGORITHMS TO ANALYZE : DECISION TREE, NAIVE BAYES CLASSIFYING ALGORITHM, RANDOM FOREST,XG BOOST AND A HYBRID CLASSIFIER. THIS IS DONE WITH THE HELP OF TRAINING DATASET CONTAINING VARIOUS FEATURES.

- JACCARD SIMILARITY MODEL USING RECENT TWEETS OF THE USER.

- COMBINING BOTH THE MODELS FOR FINAL RESULT.

Bot
Detection

Flowcharts

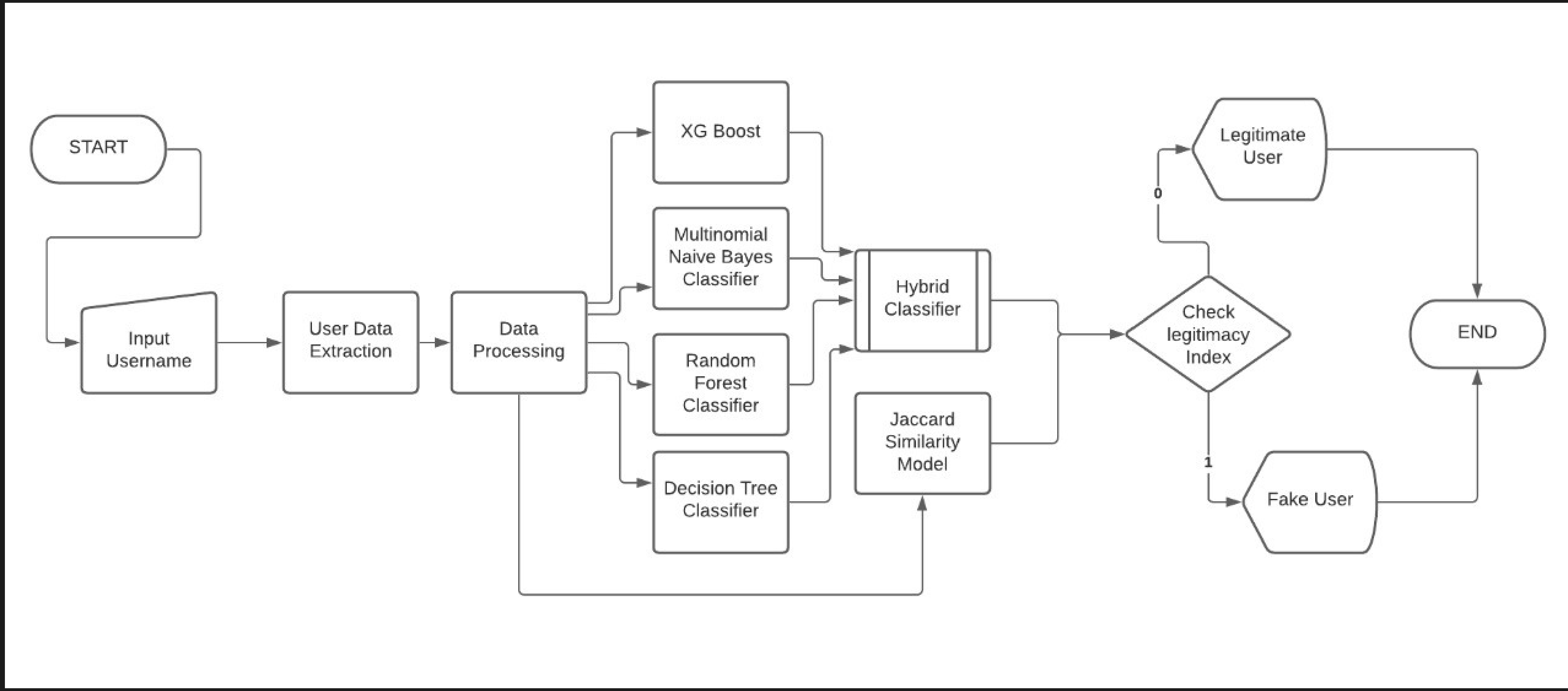


FIGURE 1: CLASSIFYING USER ACCOUNTS AS FAKE OR LEGITIMATE

Bot Detection

Data Collection

Data.csv

id	id_str	screen_na	location	descriptior	url	followers_	friends_co	listed_cou	created_at	favourites_	verified	statuses_c	lang	status	default_pr	default_pr	has_exten	name	bot
7.98E+17	"79784773	"IndyPoke.	"Indianapc	"Alerts for	"https://t.	662	3	6	"Sun Nov 1	60	FALSE	29818	"en"	{ "creat	TRUE	FALSE	FALSE	"IndyPoke.	1
4.77E+09	4.77E+09	letsplaysnake		Play with r	https://t.c	49	1	6	########	23	FALSE	9820	fr	Status(con	FALSE	FALSE	FALSE	Let's play S	1
8.31E+17	8.31E+17	NoellaKarlson		Anak Auhn	Rmbulan A	0	35	0	Mon Feb 1	58	FALSE	83	en	{"created_	TRUE	FALSE	FALSE	Noella Kar	1
8.22E+17	"82158803	"FtWorthP	"Fort Wort	"24/7 Rare	null	1148	7	12	"Wed Jan :	6	FALSE	24607	"en"	{ "creat	TRUE	FALSE	FALSE	"ft worth"	1
8.40E+17	"84033771	"ishnobop'	""	"Hi, I don't	null	2	85	0	"Fri Mar 1(	7	FALSE	11	"en"	{ "creat	TRUE	TRUE	FALSE	"Kell"	1
7.77E+17	7.77E+17	arpeggio_l	Up and do	I tweet rar	https://t.c	19	23	4	9/16/2016	2	FALSE	2768	en	{'created_	FALSE	FALSE	TRUE	Arpeggio B	1
2.16E+08	2.16E+08	IanMaksin	Chicago IL	Creating a	https://t.c	10057	7468	108	11/14/201	1905	FALSE	1459	en	Status(con	FALSE	FALSE	FALSE	Ian Maksir	1
2.89E+09	2.89E+09	imgshredder		I redact, sh	https://t.c	1464	40	157	Sun Nov 2:	996	FALSE	144584	en	{'created_	FALSE	FALSE	FALSE	Img Shredc	1
7.58E+17	7.58E+17	darndesttruisms		a bot by @	https://t.c	490	1	21	Tue Jul 26	0	FALSE	927	en	{u'contribu	TRUE	FALSE	FALSE	baby jenny	1
3.3E+09	3.3E+09	emojitoemoji		a bot by @	http://t.co	93	0	24	Tue May 2	0	FALSE	5145	en	{u'contribu	FALSE	FALSE	FALSE	[face] to [	1
7.45E+17	7.45E+17	gridgenerator		Simple generated grid		36	1	11	Tue Jun 21	0	FALSE	793	en	{u'contribu	TRUE	FALSE	FALSE	grids grids	1
2.6E+09	2.6E+09	moltar_ebooks				60	32	6	Wed Jul 02	2	FALSE	4365	en	{u'contribu	TRUE	FALSE	FALSE	MOLTAR E	1
2.39E+09	2.39E+09	DCell_pap	Mancheste	Journal paper feed fo		214	0	20	3/13/2014	0	FALSE	9727	en	Kynurenic	TRUE	FALSE	FALSE	Dendritic c	1
8.13E+17	8.13E+17	A20989664A		Free Follow for @jalfi		34	787	0	Sat Dec 24	7	FALSE	2	en	{'truncatec	TRUE	TRUE	FALSE	READ BIO	1
34716038	34716038	aaroncarte	Sony Reco	NEW AARON CARTER		571310	76070	4909	Thu Apr 23	37437	TRUE	56077	en	{u'contribu	FALSE	FALSE	TRUE	Aaron Cart	0
3013511	3013511	michellebranch		singer/son	https://t.c	292385	963	6076	Fri Mar 30	1248	TRUE	16688	en	{u'contribu	FALSE	FALSE	FALSE	Michelle B	0
56237623	56237623	stronginmyfaith				1	7	0	7/13/2009	0	FALSE	3	en	Status(con	TRUE	TRUE	FALSE	laurie linde	0
27964284	27964284	Jessicaver	5th Dimen	@TheVero	https://t.c	222659	352	3261	Tue Mar 3:	143	TRUE	10999	en	{u'contribu	FALSE	FALSE	FALSE	Jessica Ver	0
5.53E+08	5.53E+08	resargentc	brasil	meio ogra mas o cora		646	446	2	4/13/2012	13440	FALSE	25292	pt	null	FALSE	FALSE	FALSE	rebosta	0
3.81E+09	3.81E+09	crazyl1f				2	5	0	Tue Sep 29	0	FALSE	2	zh-cn	{'truncatec	TRUE	TRUE	FALSE	Songgaoyu	0
1.53E+09	1.53E+09	Taniasimo	Milano	Mechanical engineer,		27	25	0	6/19/2013	17	FALSE	50	it	Status(in_r	TRUE	FALSE	FALSE	Tania Simc	0
1.29E+09	1.29E+09	YCPRProf	ëÔë£ë_-%	We are the	https://t.c	316	192	9	3/22/2013	2257	FALSE	7329	en	Status(in_r	FALSE	FALSE	TRUE	Dr. K. McB	0
19980906	19980906	bandofhorses		'Why Are Y	https://t.c	211616	8617	4071	Tue Feb 03	1941	TRUE	2037	en	{u'contribu	FALSE	FALSE	FALSE	Band of Hc	0
3.78E+08	3.78E+08	sparker	SF / LA / N	Napster, P	https://t.c	429604	600	4999	Thu Sep 22	30	TRUE	497	en	{'created_	FALSE	FALSE	FALSE	Sean Parke	0

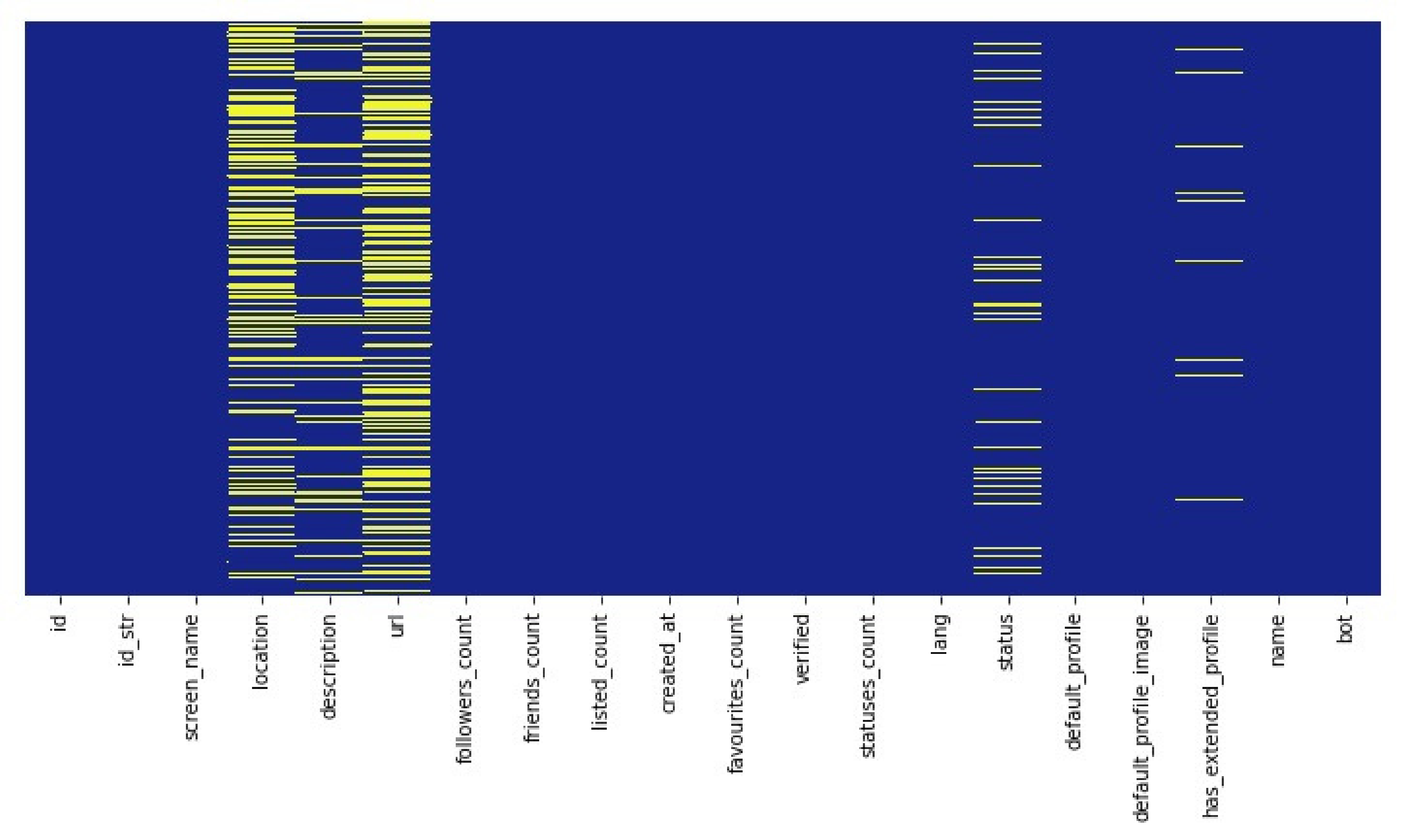
Data Preprocessing

Technologies used:

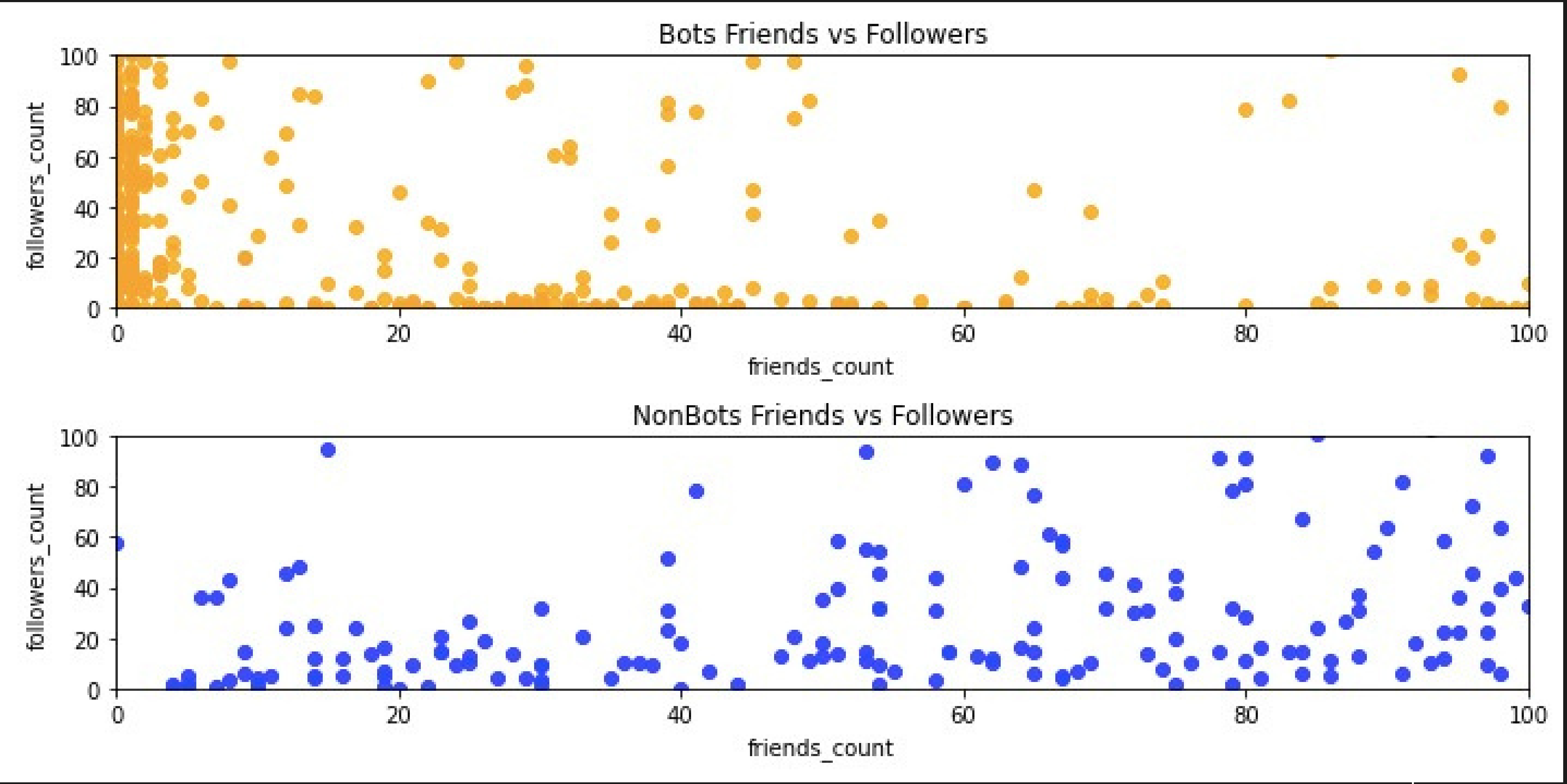
-Twitter API
-Tweepy library
-get_user() method
-Anaconda
-Jupyter Notebook
-Python3 Libraries

Algorithm :

- Identifying missing values and imbalance in data.
- Feature extraction
- Feature engineering
- Discarding unwanted attributes



Heatmap of training data


Bots Friends vs Followers
followers_count
friends_count
0
20
40
60
80
100
NonBots Friends vs Followers
followers_count
friends_count
0
20
40
60
80
100

Features used for training the machine learning model :

- Screen name
- Name
- Description
- Status
- Verified
- Followers_count
- Friends_count
- Statuses_count
- Listed_Count

Cleaning of Data

Attributes containing strings and tweets extracted from the user accounts needed to be cleaned.

Steps performed for cleaning -

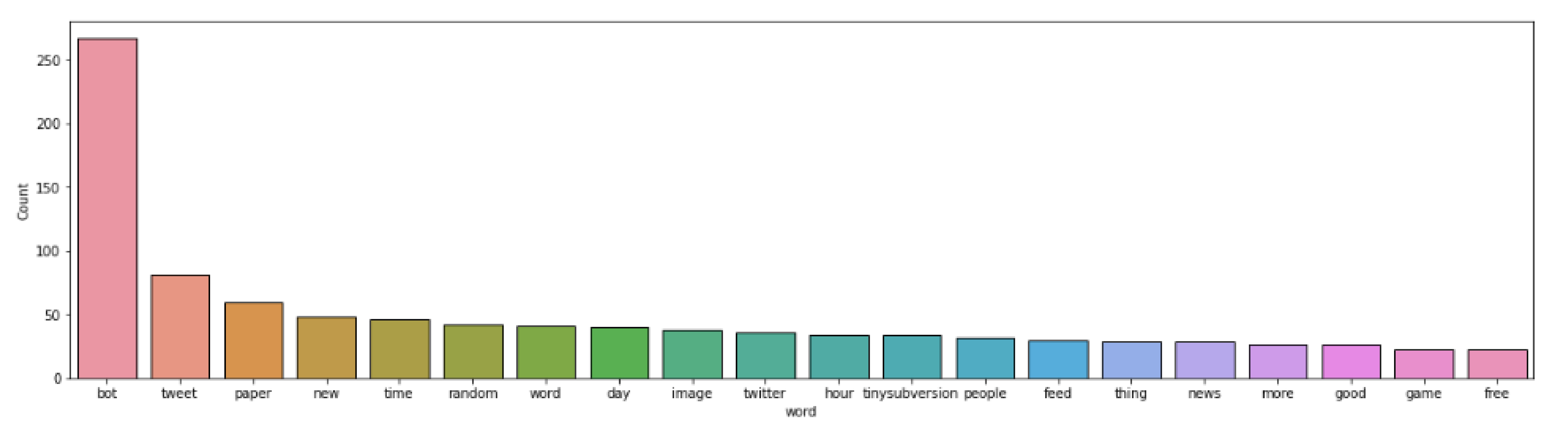
- Removing links, @mentions and two & less letter words
- Lemmatization
- Tokenization
- Stop word removal
- Conversion to lowercase

Atributes like status, name, screen name and description were converted into binary attributes.

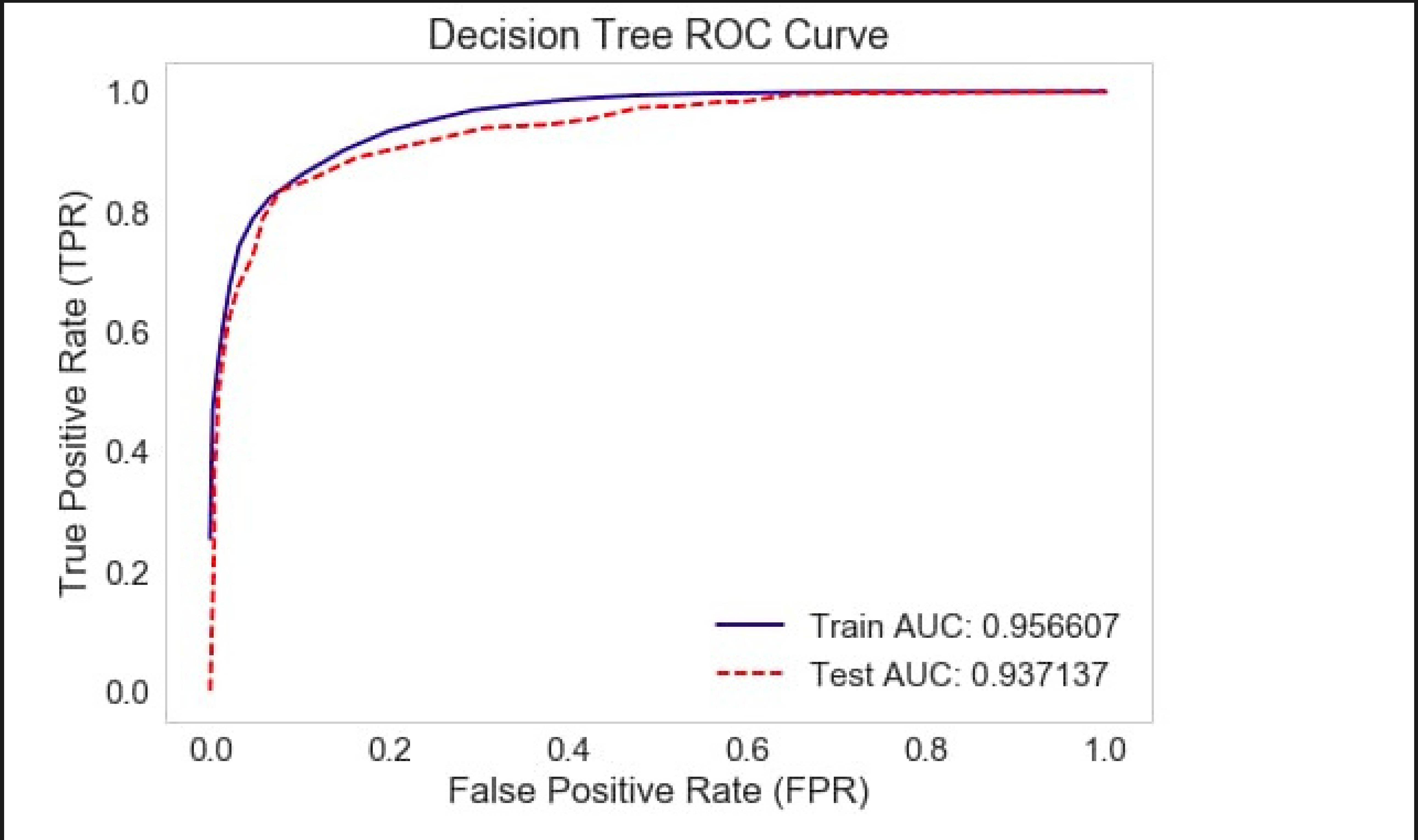
For this a bag of words was created using 50 most frequent words from each attribute.

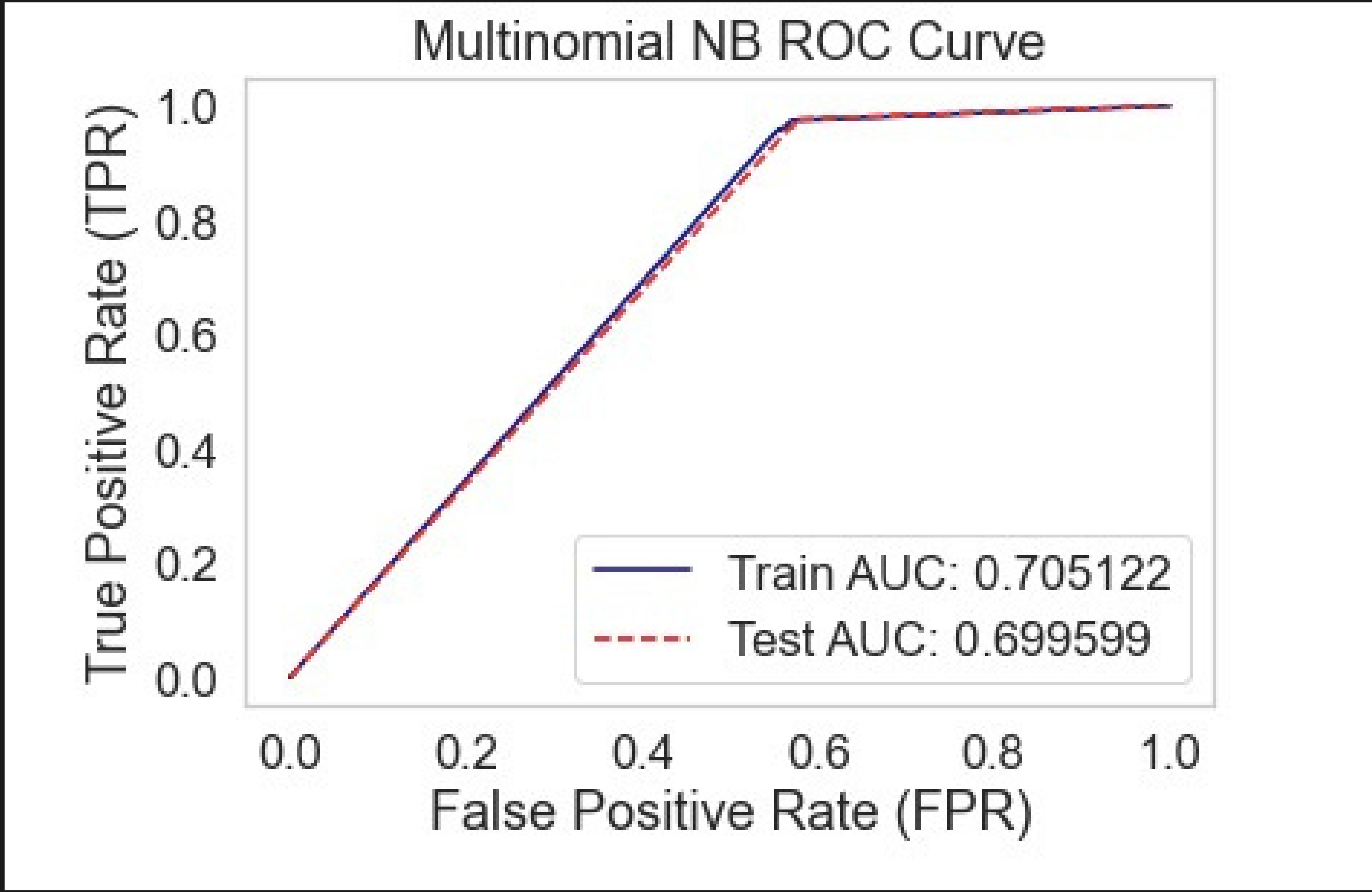
If the given attributes contained words from the bag of words, they were labelled as 1 otherwise 0.

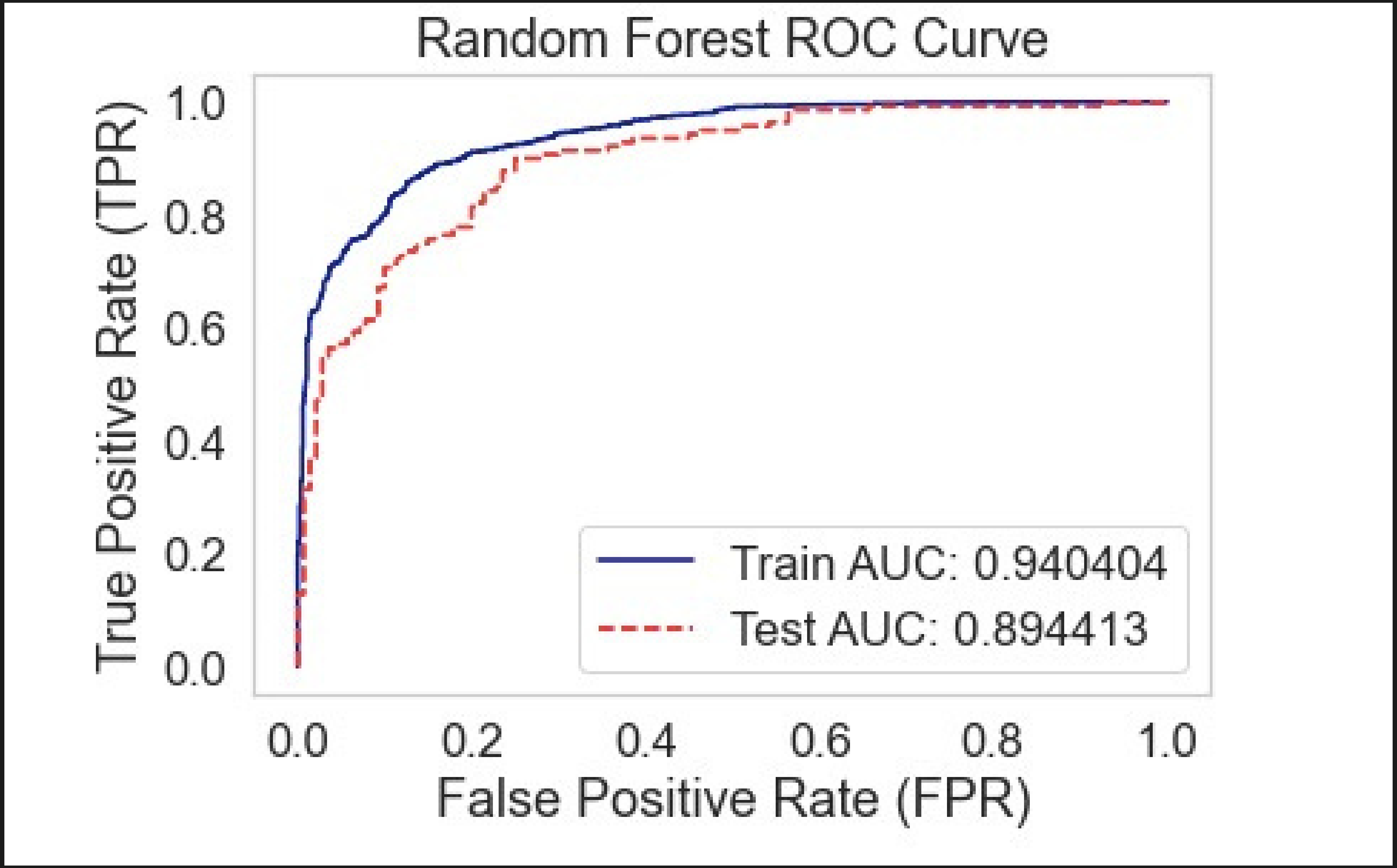
The below figure contains the 20 most frequent words found in description of bot accounts from the training dataset.

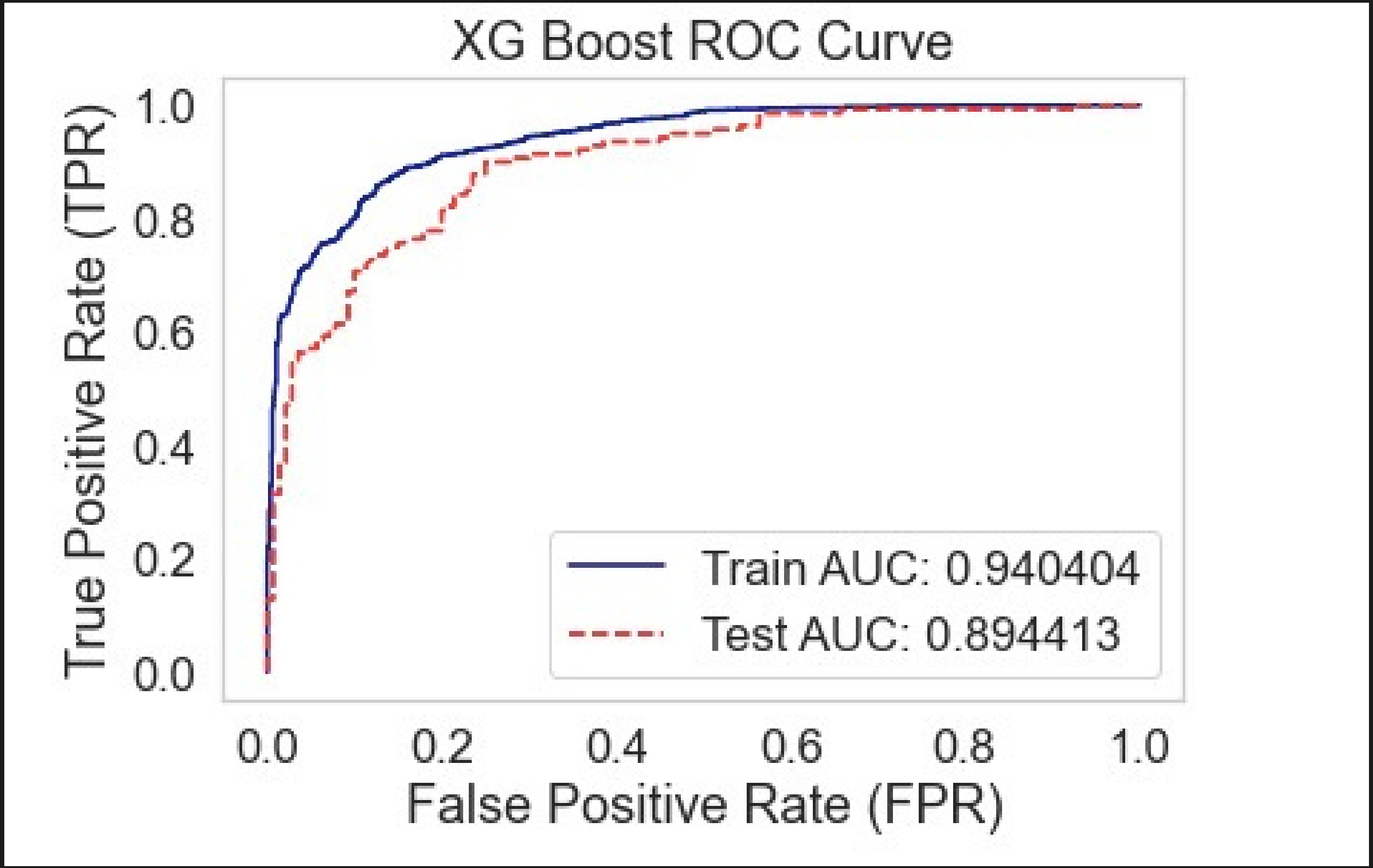


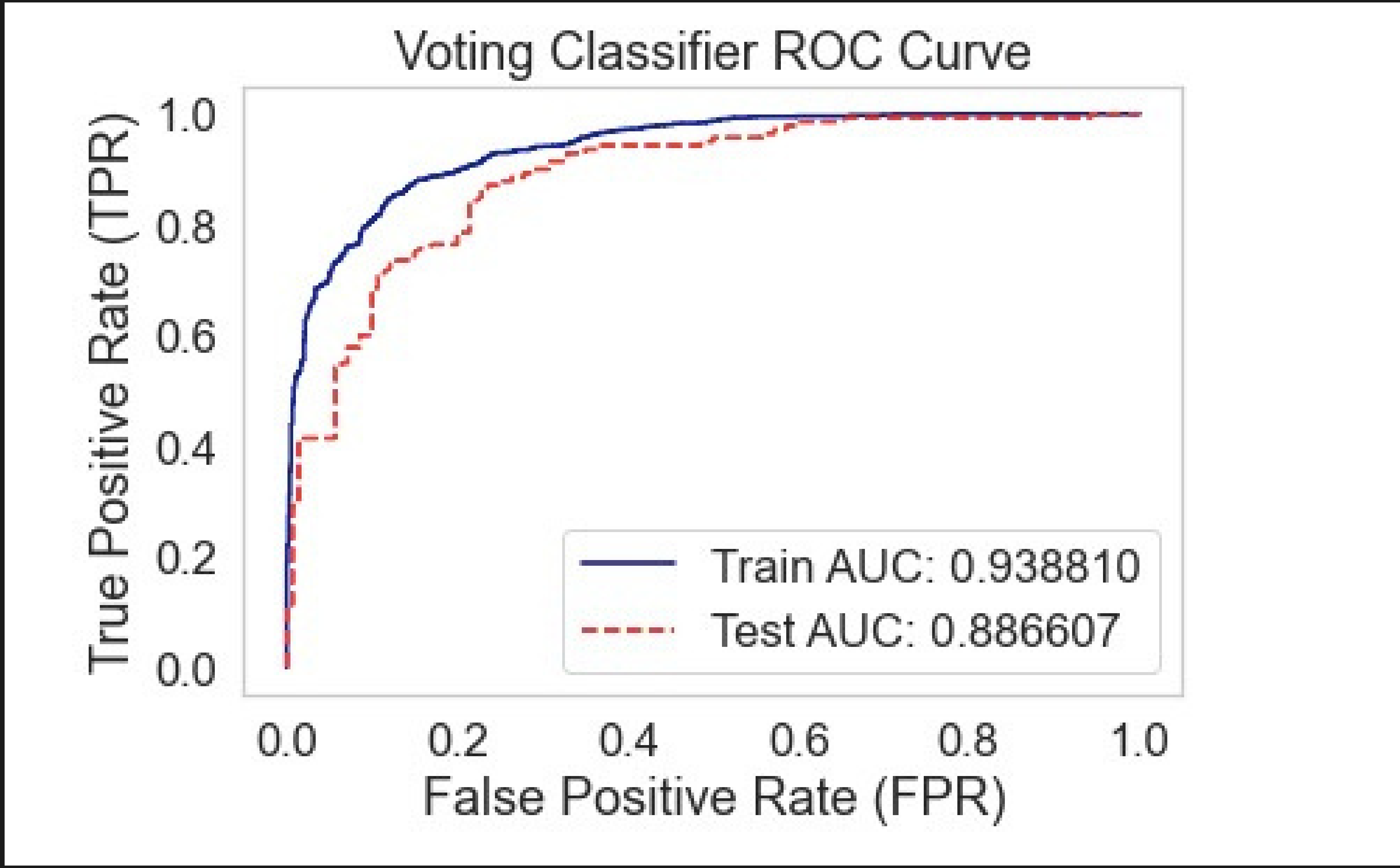
Implementation




Multinomial NB ROC Curve
True Positive Rate (TPR)
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate (FPR)
Train AUC: 0.705122
Test AUC: 0.699599


Random Forest ROC Curve
True Positive Rate (TPR)
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate (FPR)
Train AUC: 0.940404
Test AUC: 0.894413


XG Boost ROC Curve
True Positive Rate (TPR)
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate (FPR)
Train AUC: 0.940404
Test AUC: 0.894413



Voting Classifier ROC Curve
True Positive Rate (TPR)
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate (FPR)
Train AUC: 0.938810
Test AUC: 0.886607

Model	Training Accuracy	Testing Accuracy	Training Precision	Testing Precision	Training Data Recall	Testing Data Recall
Decision Tree Classifier	88.20%	87.85%	90.50%	91.10%	84.40%	83.60%
Multinomial NB Classifier	67.80%	69.70%	59.30%	62.50%	96.20%	97.10%
Random Forest Classifier	84.80%	84.40%	86.50%	87.60%	79.60%	79.80%
XGB Classifier	98.80%	83.50%	99.20%	84.50%	98.20%	82.10%
Hybrid Ensemble Model	92.30%	90.00%	91.10%	88.80%	92.70%	91.40%

Jaccard Similarity model

HERE, TO FIND SIMILARITY BETWEEN THE TWEETS, JACCARD SIMILARITY IS USED. IN THIS SIMILARITY TEST, THE EXPERIMENT DEPICTED THE JACCARD SIMILARITY SCORE BETWEEN THE 100 MOST RECENT TWEETS OF A GIV-EN TWITTER ACCOUNT. BEFORE THE SIMILARITY SCORE WAS COMPUTED, THE TWEETS MADE AVAILABLE FROM THE DATASET WERE IN THEIR RAW FORM AND HAD TO BE PREPROCESSED FIRST FOR EFFECTIVE USE. FURTHER, THE TOKENIZING OF THE TWEETS PERFORMED USING REGEXP TOKENIZER AND LEMMATIZED THE TWEETS USING WORDNET LEMMATIZER FROM NLTK LIBRARY. AFTER THIS, THE JACCARD SIMILARITY SCORE WAS COMPUTED. THE ACCURACY UP TO 93.2% IS ACHIEVED AFTER COMBINING THE RESULTS FROM BOTH THE HYBRID MODEL AND THE SIMILARITY MODEL.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Conclusion

- PRESENCE OF ARTIFICIAL BOT ACCOUNTS ON THE SOCIAL MEDIA PLATFORM POSE THREAT TO THE PRIVACY OF THE LEGITIMATE USERS.
- IN THE PROPOSED MODEL VARIOUS MACHINE LEARNING ALGORITHMS WERE USED AS WEAK LEARNERS TO MAKE A HYBRID MODEL THAT COULD SUCCESSFULLY CLASSIFY 90% OF THE ACCOUNTS USING THE PREPROCESSED TWITTER DATA.
- THE PROJECT ALSO CONSISTS OF CLASSIFICATION BASED ON JACCARD SIMILARITY MODEL WHICH USES RECENT TWEETS POSTED BY THE USERS.

The Research paper was published as a chapter in the following book.

Book Name - HANDBOOK OF RESEARCH ON DATA PREPROCESSING, ACTIVE LEARNING, AND COST PERCEPTIVE APPROACHES FOR RESOLVING DATA IMBALANCE.

Publication - IGI GLOBAL

LINK - https://www.igi-global.com/submission/book-project-chapters/?projectid=03dcf765-2e75-4bf4-9089-305370caa331

Future Work

SOME METHODS, SUCH AS GUIDED LEARNING APPROACHES, WERE EXTENSIVELY DISCUSSED IN THIS WORK. TO COMPREHEND, REINFORCE OR DISCOVER NEW RESULTS, SEVERAL APPROACHES REQUIRE MORE EXPLORATION. IT STIMULATES THE RESEARCH COMMUNITY TO DISCOVER NEW APPROACHES AND IM□PROVE EXISTING APPROACHES. THE ABOVE MODEL CAN BE UTILIZED FOR REAL TIME APPLICATIONS. WITH THE AWARENESS OF BOT ACCOUNTS AMONG USERS, IT WOULD BECOME CONVENIENT AND HIGHLY LUCRATIVE FOR PEOPLE AND ORGANIZATIONS TO DETECT THEM ON OSNS. IN FUTURE, THIS MODEL WILL BE COMPARED WITH OTHER AVAILABLE TECHNIQUES AND BY INCLUDING THE USAGE OF NETWORK INFORMATION OF USERS.

References

[**1**] LEYLA BILGE ET AL. “ALL YOUR CONTACTS ARE BELONG TO US: AUTOMATED IDENTITY THEFT ATTACKS ON SOCIAL NETWORKS”. IN: PROCEEDINGS OF THE 18TH INTERNATIONAL CONFERENCE ON WORLD WIDE WEB. WWW ’09. MADRID, SPAIN: ASSOCIATION FOR COMPUTING MACHINERY, 2009, PP. 551–560. ISBN: 9781605584874. DOI: 10.1145/1526709.1526784. URL: HTTPS://DOI.ORG/10.1145/1526709.1526784.

[**2**] LEI JIN, HASSAN TAKABI, AND JAMES B.D. JOSHI. “TOWARDS ACTIVE DETECTION OF IDENTITY CLONE ATTACKS ON ONLINE SOCIAL NETWORKS”. IN: PROCEEDINGS OF THE FIRST ACM CONFERENCE ON DATA AND APPLICATION SECURITY AND PRIVACY. CODASPY ’11. SAN ANTONIO, TX, USA: ASSOCIATION FOR COMPUTING MACHINERY, 2011, PP. 27–38. ISBN: 9781450304665. DOI: 10 . 1145 / 1943513 . 1943520. URL: HTTPS://DOI.ORG/10.1145/1943513.1943520.

[**3**] G. KONTAXIS ET AL. “DETECTING SOCIAL NETWORK PROFILE CLONING”. IN: 2011 IEEE INTERNATIONAL CONFERENCE ON PERVASIVE COMPUTING AND COMMUNICATIONS WORKSHOPS (PERCOM WORKSHOPS). 2011, PP. 295–300.

[**4**] JENNIFER GOLBECK. “BENFORD’S LAW APPLIES TO ONLINE SOCIAL NETWORKS”. IN: PLOS ONE 10.8 (AUG. 2015), PP. 1–10. DOI: 10 . 1371 / JOURNAL . PONE . 0135169. URL: HTTPS://DOI.ORG/10.1371/JOURNAL.PONE.0135169.

[**5**] NIKAN CHAVOSHI, HOSSEIN HAMOONI, AND ABDULLAH MUEEN. “DEBOT: TWITTER BOT DETECTION VIA WARPED CORRELATION.” IN: ICDM. 2016, PP. 817–822.

References

[6] CLAUDIA MEDA ET AL. “SPAM DETECTION OF TWITTER TRAFFIC: A FRAMEWORK BASED ON RANDOM FORESTS AND NON-UNIFORM FEATURE SAMPLING”. IN: 2016 IEEE/ACM INTERNATIONAL CONFERENCE ON ADVANCES IN SOCIAL NETWORKS ANALYSIS AND MINING (ASONAM). IEEE. 2016, PP. 811–817.

[7] BUKET ER¸SAHIN ET AL. “TWITTER FAKE ACCOUNT DETECTION”. IN: 2017 INTERNATIONAL CONFERENCE ON COMPUTER SCIENCE AND ENGINEERING (UBMK). IEEE. 2017, PP. 388–392.

[8] SAGAR GHARGE AND MANIK CHAVAN. “AN INTEGRATED APPROACH FOR MALICIOUS TWEETS DETECTION USING NLP”. IN: 2017 INTERNATIONAL CONFERENCE ON INVENTIVE COMMUNICATION AND COMPUTATIONAL TECHNOLOGIES (ICICCT). IEEE. 2017, PP. 435–438.

[9] MOHAMMADREZA MOHAMMADREZAEI, MOHAMMAD EBRAHIM SHIRI, AND AMIR MASOUD RAHMANI. “IDENTIFYING FAKE ACCOUNTS ON SOCIAL NETWORKS BASED ON GRAPH ANALYSIS AND CLASSIFICATION ALGORITHMS”. IN: SECURITY AND COMMUNICATION NETWORKS 2018 (2018).

[10] JENNIFER GOLBECK. “BENFORD’S LAW CAN DETECT MALICIOUS SOCIAL BOTS”. IN: FIRST MONDAY 24.8 (AUG. 2019). DOI: 10.5210/FM.V24I8.10163. URL: HTTPS://JOURNALS.UIC.EDU/OJS/INDEX.PHP/FM/ARTICLE/VIEW/10163.

References

[11] S. GHARGE AND M. CHAVAN, “AN INTEGRATED APPROACH FOR MALICIOUS TWEETS DETEC- TION USING NLP,” IN 2017 INTERNATIONAL CONFERENCE ON INVENTIVE COMMUNICATION AND COMPUTATIONAL TECHNOLOGIES (ICICCT), IEEE, 2017, PP. 435–438.

[12] O. VAROL, E. FERRARA, C. DAVIS, F. MENCZER, AND A. FLAMMINI, “ONLINE HUMAN- BOT INTERACTIONS: DETECTION, ESTIMATION, AND CHARACTERIZATION,” IN PROCEEDINGS OF THE INTERNATIONAL AAAI CONFERENCE ON WEB AND SOCIAL MEDIA, VOL. 11, 2017.

[13] S. KUDUGUNTA AND E. FERRARA, “DEEP NEURAL NETWORKS FOR BOT DETECTION,” INFORMA- TION SCIENCES, VOL. 467, PP. 312–322, 2018.

[14] M. MOHAMMADREZAEI, M. E. SHIRI, AND A. M. RAHMANI, “IDENTIFYING FAKE ACCOUNTS ON SOCIAL NETWORKS BASED ON GRAPH ANALYSIS AND CLASSIFICATION ALGORITHMS,” SECURITY AND COMMUNICATION NETWORKS, VOL. 2018, 2018.

[15] J. GOLBECK, “BENFORD’S LAW CAN DETECT MALICIOUS SOCIAL BOTS,” FIRST MONDAY, VOL. 24, NO. 8, AUG. 2019. DOI: 10.5210/FM.V24I8.10163. [ONLINE]. AVAILABLE: HTTPS: //JOURNALS.UIC.EDU/OJS/INDEX.PHP/FM/ARTICLE/VIEW/10163.

Acknowledgement

WE TAKE THIS OPPORTUNITY TO EXPRESS HEARTFELT GRATITUDE TO OUR PROJECT GUIDE, DR. DIPTI P. RANA, ASSISTANT PROFESSOR IN COMPUTER ENGINEERING DEPARTMENT, SVNIT SURAT FOR HER VALUABLE GUID- ANCE, CONSTANT ENCOURAGEMENT, AND HELPFUL FEEDBACK ALL THROUGHOUT VARIOUS STAGES OF WORK.
WE WOULD ALSO LIKE TO THANK OUR HEAD OF DEPARTMENT, DR. M. A. ZAVERI, COMPUTER EN- GINEERING DEPARTMENT FOR ALL THE SUPPORT. WE ARE VERY GRATEFUL TO SVNIT SURAT AND ITS STAFF FOR PROVIDING US WITH THIS OPPORTUNITY WHICH AIDED US IN ACQUIRING REQUIRED KNOWLEDGE TO BE SUCCESSFUL IN OUR WORK.