



NLP - TOXIC COMMENT CLASSIFICATION

AGENDA

Business Case

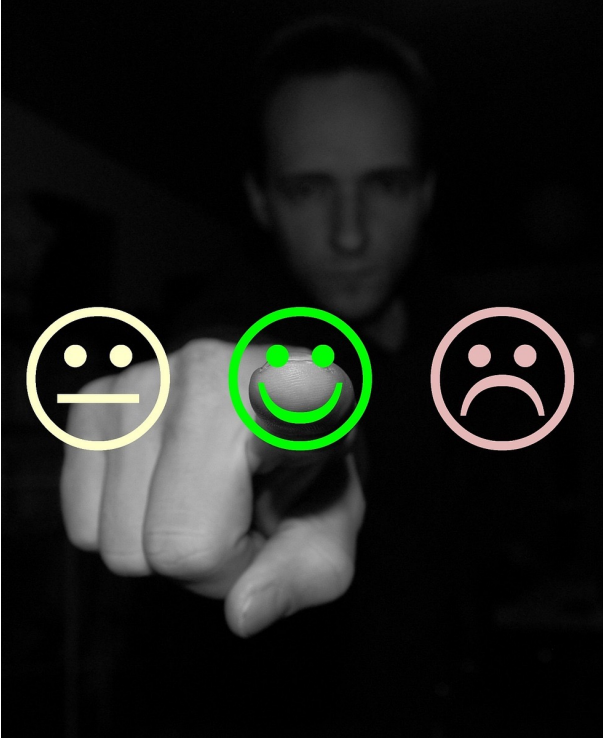
Data

Modeling

Best Model

Limitations

BUSINESS CASE



Monitoring Online Discussion Forum

- Company XYZ operating an online discussion forum
- New to market
- Inexperienced in content moderation
- Limited resources for content moderation

DATA

[Kaggle Toxic Comment Classification Challenge](#)

Wikipedia comments from users

Normal

144,277

Toxic

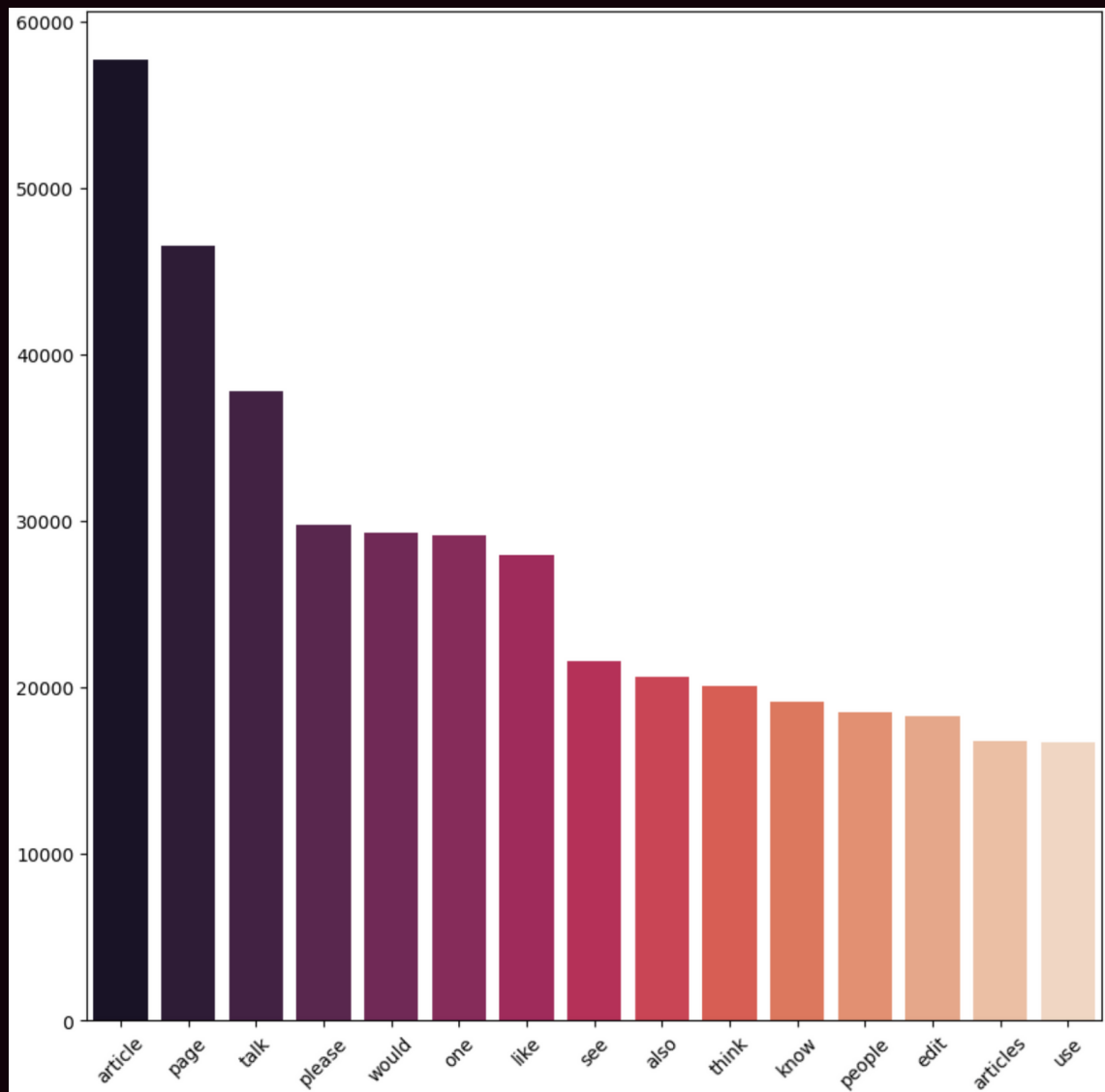
15,294

Total

159,571



Top 20 Words



TOP 20 TOXIC WORDS



Toxic Comments

WHOEVER PUT A
NOTICES ON MY PAGE. I
WILL KILL U

A

STUPID PEACE OF SHIT
STOP DELETING MY
STUFF ASSHOLE GO DIE
AND FALL IN A HOLE GO
TO HELL!

B

BLA BLA BLA....SUCK IT
IRISHGUY =)

C

Normal Comments

I AM MOST CERTAINLY
NOT A SOCKPUPPET OF
THECHIEF, PLEASE
RETRACT YOUR
FRADULENT CLAIM AT
ONCE.

X

YOU, SIR, ARE MY HERO.
ANY CHANCE YOU
REMEMBER WHAT PAGE
THAT'S ON?

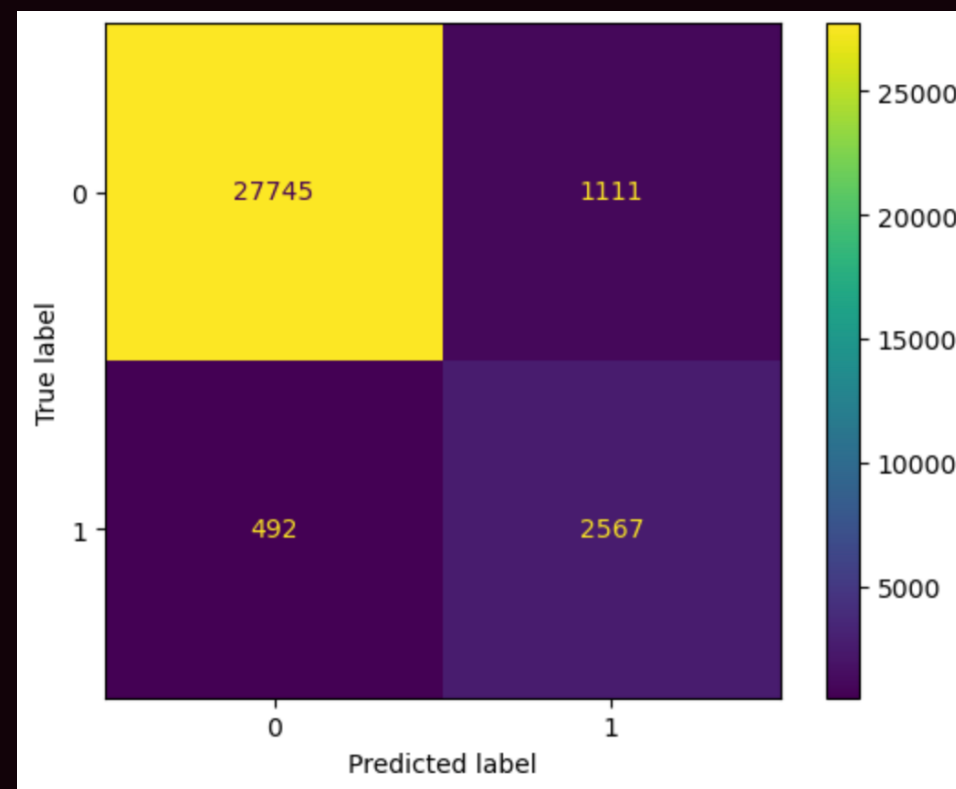
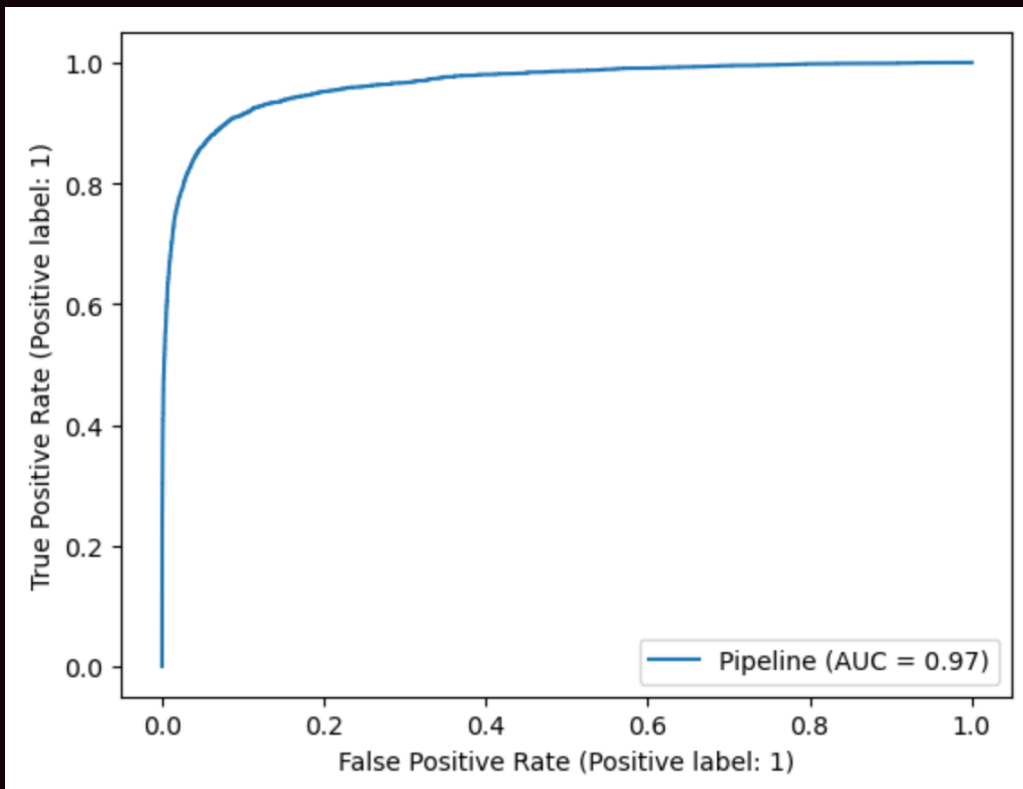
Y

- NEILN'S BANG ON -
YOU AREN'T BEING
HARASSED AT ALL, IF
YOU DON'T EDIT
DISRUPTIVELY YOU
DON'T GET REVERTED.

Z

Models

Models	Recall	F1	AUC	FN	FP
Model1 - MultinomialNB - CountVectorizer	83	66	0.89	2495	5
Model2 - MultinomialNB - CountVectorizer, SMOTE	83	66	0.95	557	1869
Model2b - MultinomialNB - best params - CountVectorizer - SMOTE	83	66	0.95	512	2156
Model3- MultinomialNB -best params - TfidfVectorizer - SMOTE	82	63	0.94	561	2403
Model4 - RandomForest - TfidfVectorizer - SMOTE	60	69	0.96	1231	390
Model4b - RandomForest - CountVectorizer - SMOTE	59	69	0.96	1257	349
Model5 - RandomForest - best params - CountVectorizer - SMOTE	82	31	0.83	554	10000
Model5b - RandomForest - best params - CountVectorizer - RandomOverSampler	86	31	0.86	416	11432
Model6 - XGBoost - CountVectorizer - RandomOverSampler	79	79	0.95	649	1058
Model6 - XGBoost - best params - CountVectorizer - RandomOverSampler	82	75	0.96	547	1110
Model7 - Ensemble - Stacking - MultinomialNB - XGBoost -	84	76	97	492	1111



Limitations

“ NAZI FILTH” IS IMPOLITE ”

X

“ YOU ARE ONE SAD LITTLE INDIVIDUAL. ”

Y

“
ANTI-SEMITIC CRAP
THE WHOLE ARTICLE IS AN ANTI-SEMITIC
CONSPIRACY THEORY CHERISHED BY
ARYAN-SUPREMACIST RIGHTISTS! ”

Z

“
YOU DISGRACE TO HUMANITY. STOP WASTING
VALUABLE INTERNET RESOURCES (YOUR
KILLING THE FUCKING PLANET) AND THROW
OUT YOUR PC. PS IT'S LIKE 3AM WHAT THE FARK
KINDA LIFE DO YOU HAVE ON WIKIPEDIA 24/7.
A VERY SAD AND LONELY ONE LOL. LIFE'S TOO
SHORT FIND SOMETHING BETTER TO DO ”

W

THANK YOU

