

Social Voice on Smart Watch





Problem Statement

- To find what reddit user's social voice
 - User Experience
 - Smart Watch features
 - Technical Problems
 - Series of these smart watches
 - Apple Watch 3
 - Garmin Fenix 3

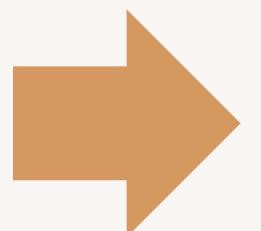
For example, Garmin's battery life is more longer than Apple Watch Series 4

Data Processing

Collect Data

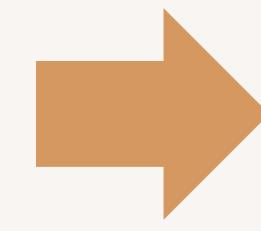


*r/Garmin
r/AppleWatch*

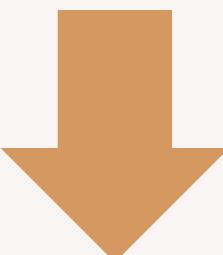


Clean Data

*Drop Duplicate
Drop Punctuation
Drop Null*

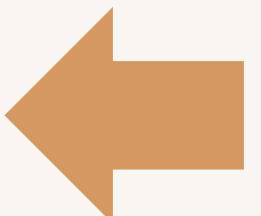


Concatenate Data

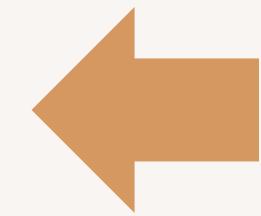


Explore Data

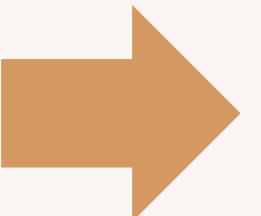
Prediction



Train/Test Split



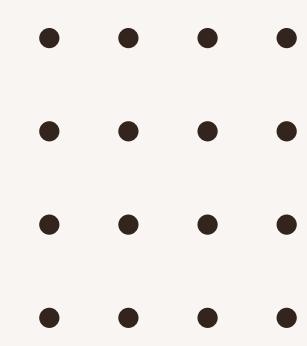
Modeling



Evaluate

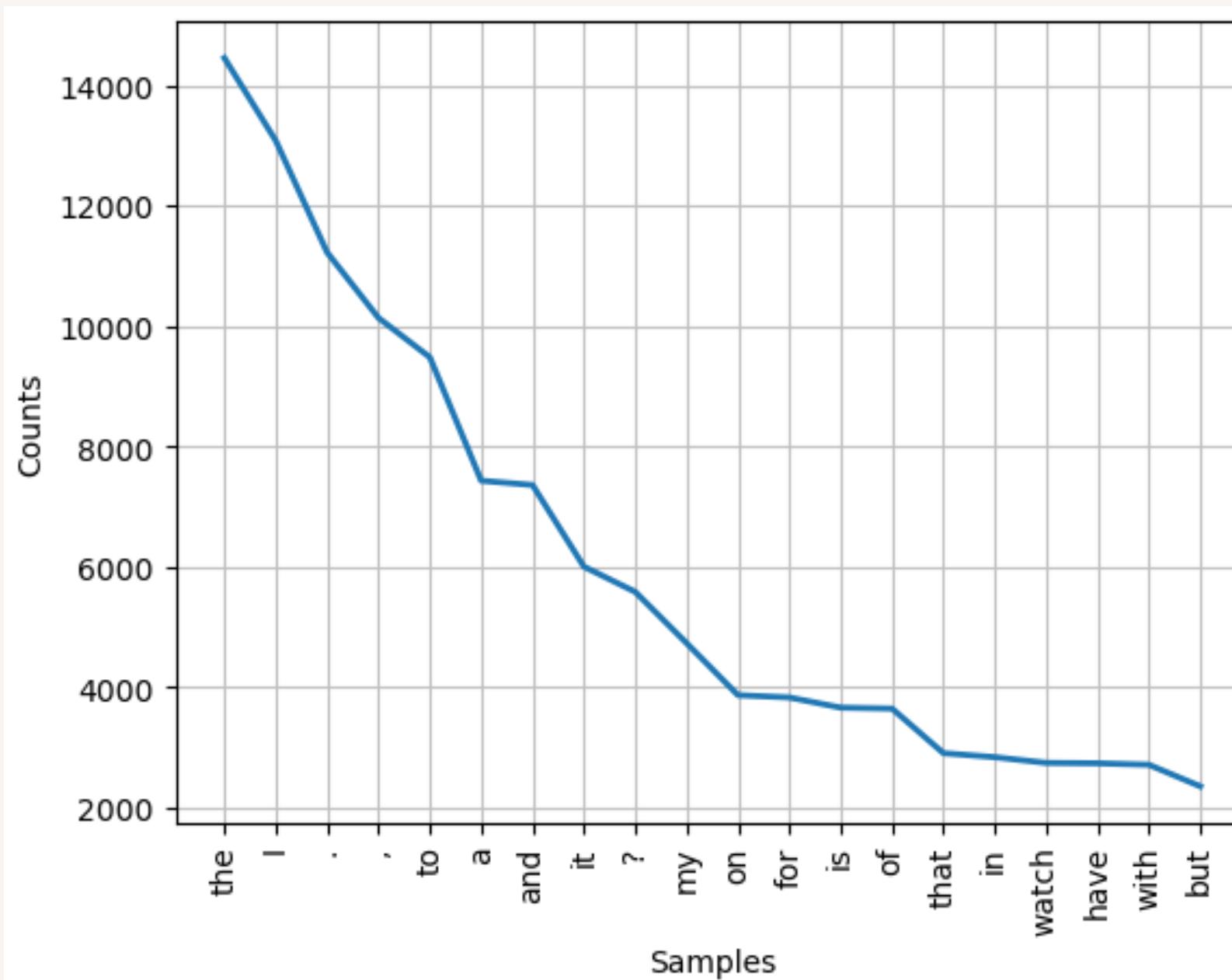
• • •
• • •
• • •
• • •

Exploratory Data Analysis

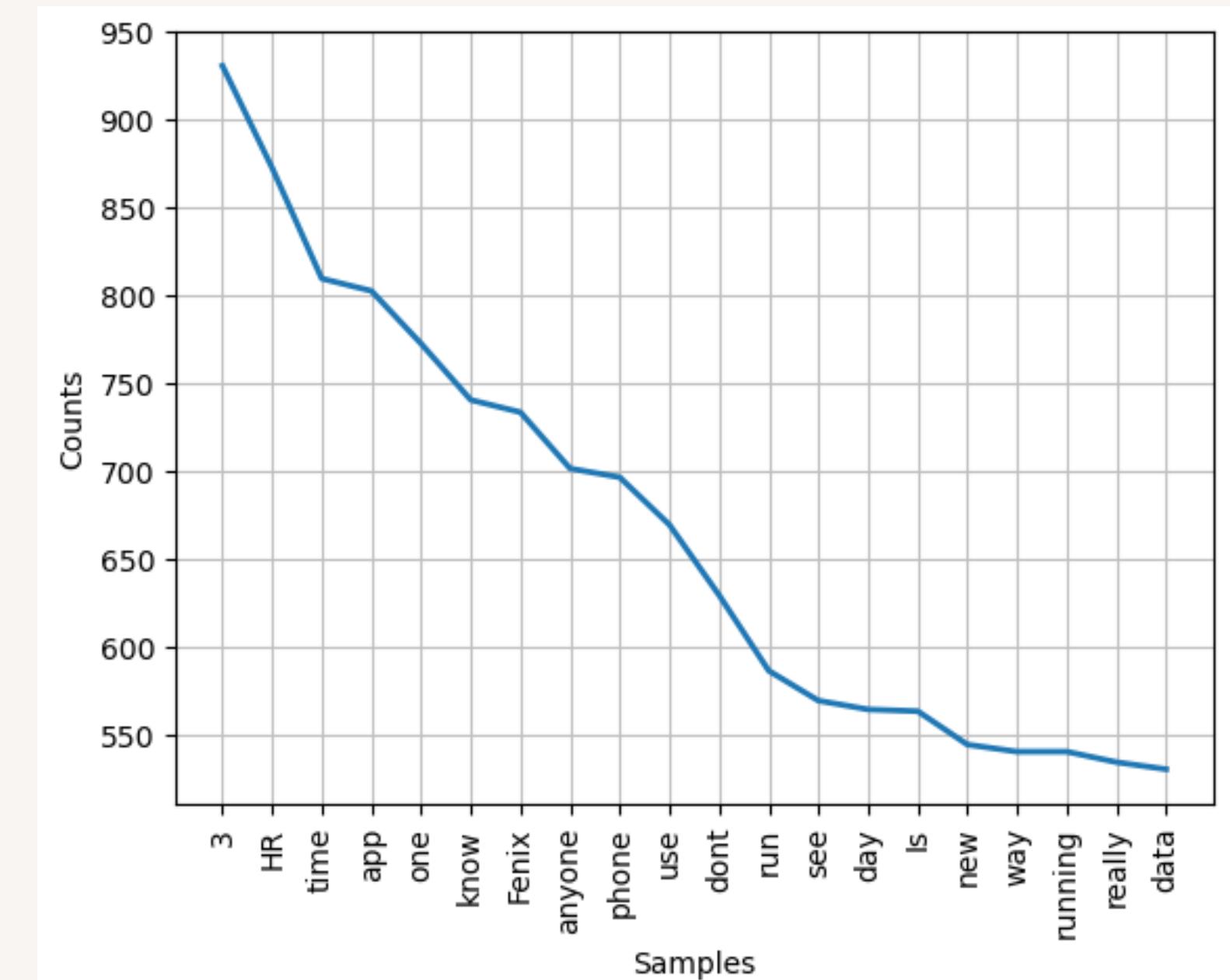


Distribution of classes: Garmin(64%) and AppleWatch(35%)

Before vectorize



After vectorize



• • • •
• • • •
• • • •
• • • •

Modeling and Evaluation

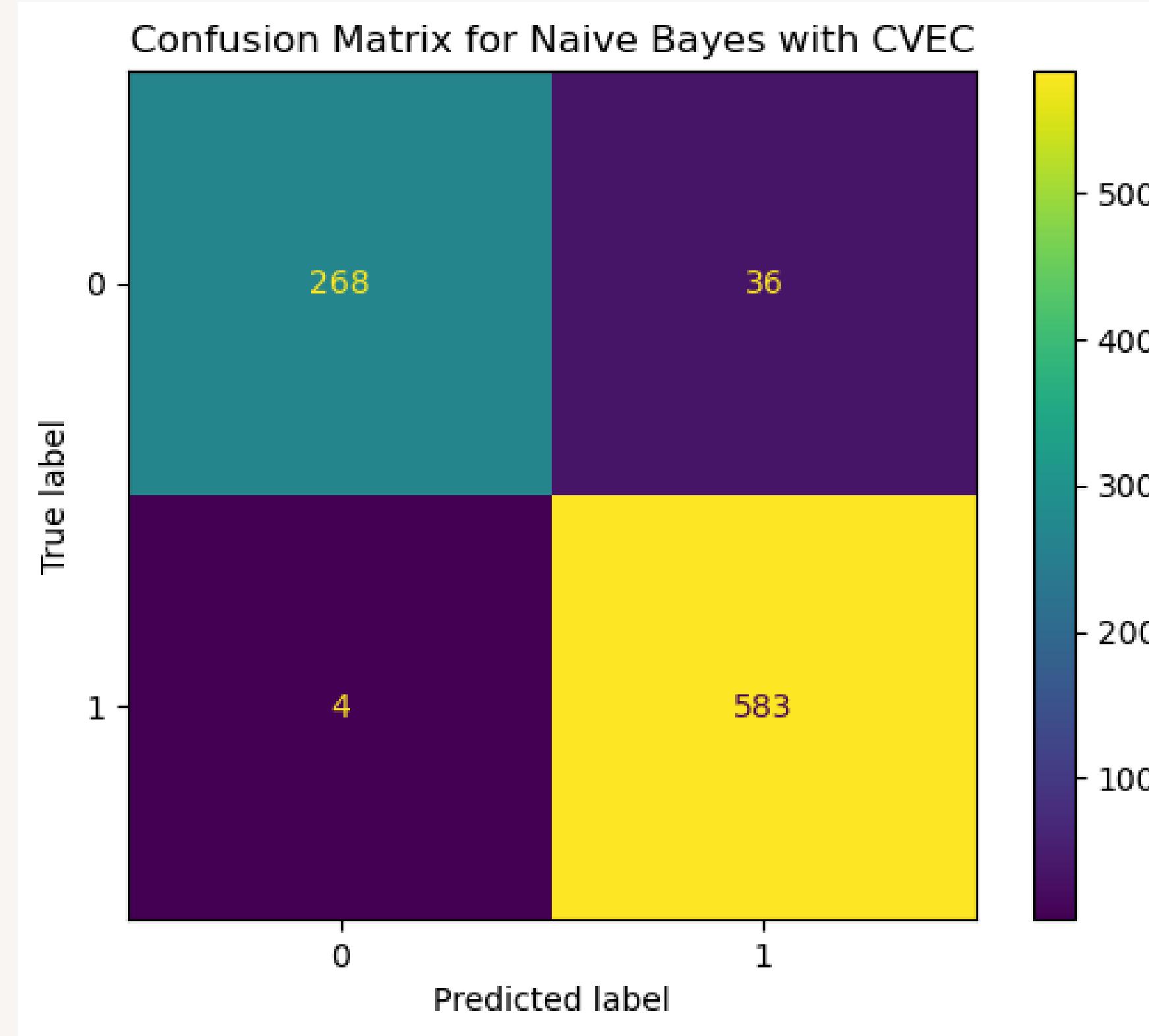
Before modeling, grid search with cross-validation

- Naive Bayes (CVEC/TVEC)
- Random Forest(CVEC/TVEC)
- Logistic Regression (CVEC/TVEC)
- Adaboost (CVEC/TVEC)
- SVC(CVEC/TVEC)



• • • •
• • • •
• • • •
• • • •

Naive Bayes with CVEC



True Positive: 583

True Negative: 268

False Positive: 36

False Negative: 4

Train Score: 0.98

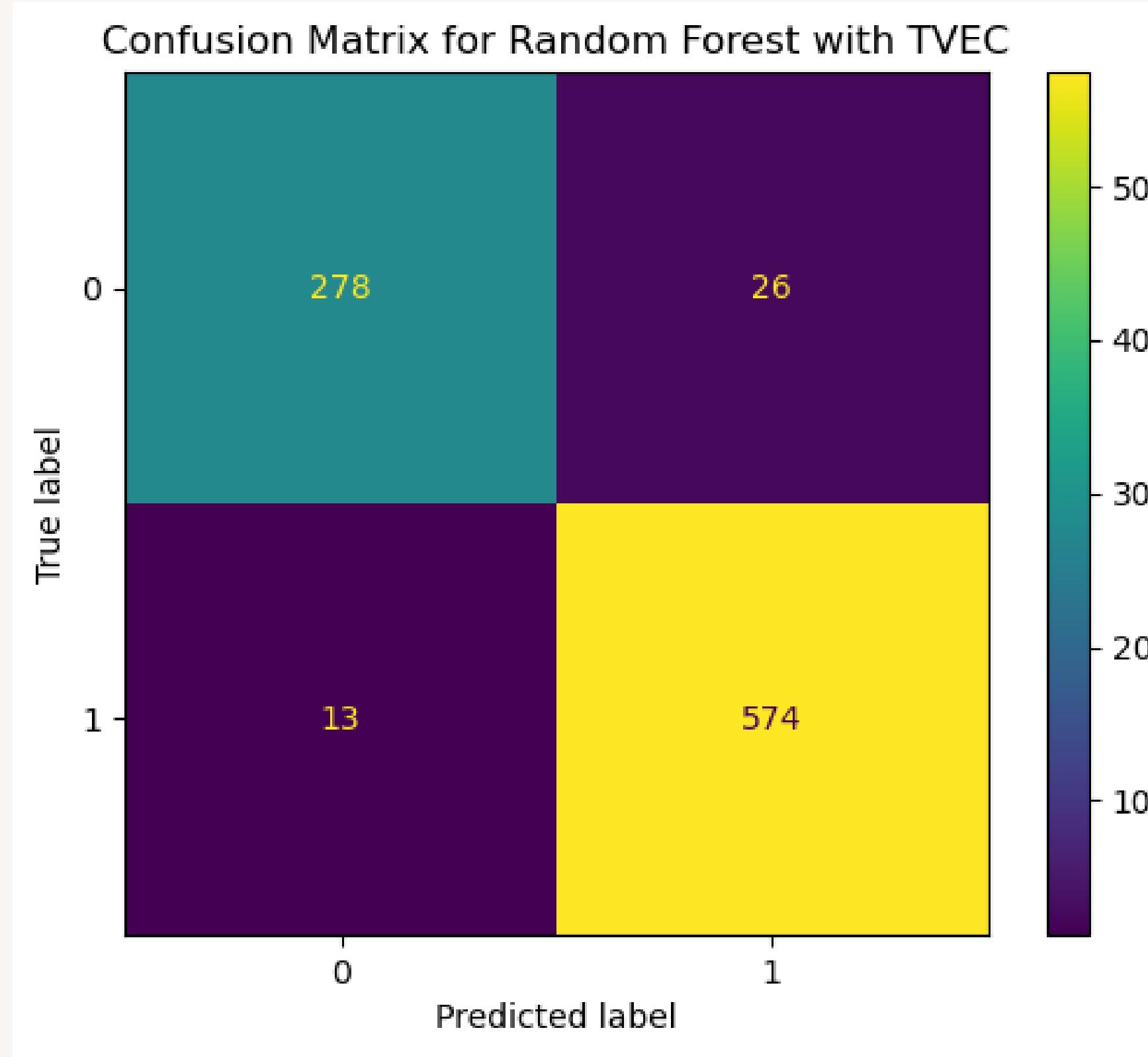
Test Score: 0.94

Precision: 0.96

Recall: 0.96

F1: 0.95

Random Forest with TVEC



True Positive: 574

True Negative: 278

False Positive: 26

False Negative: 13

Train Score: 1.0

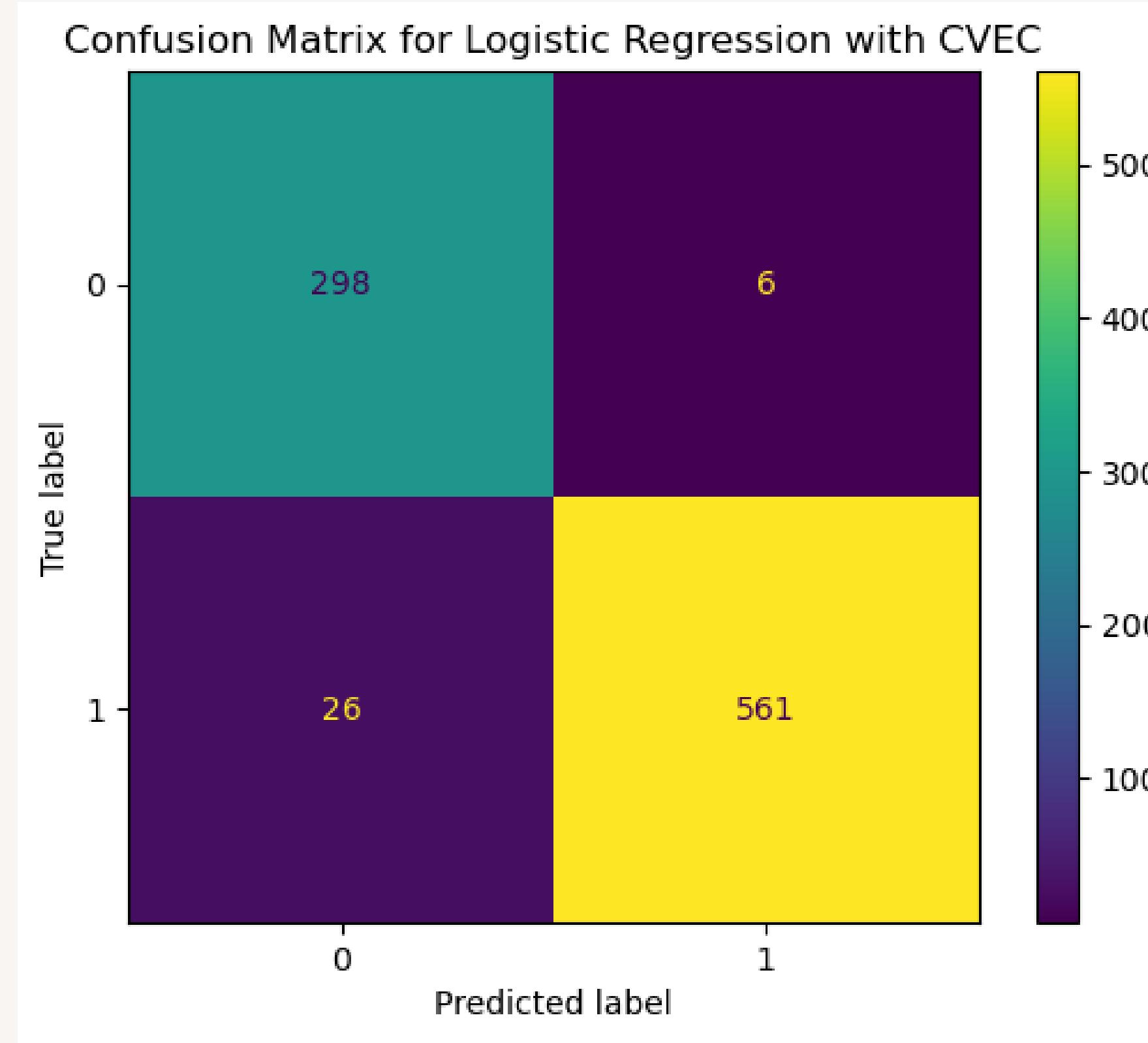
Test Score: 0.95

Precision: 0.96

Recall: 0.96

F1: 0.96

Logistic Regression with CVEC (Best Model)



True Positive: 574

True Negative: 278

False Positive: 26

False Negative: 13

Train Score: 0.99

Test Score: 0.96

Precision: 0.96

Recall: 0.96

F1: 0.96

Additional Models

Model Name	Best Parameter	Train Score	Test Score	Precision	Recall	F1
Naive Bayes (TVEC)	alpha: 0.1	0.97	0.93	0.94	0.91	0.93
Random Forest (CVEC)	max_depth: None, n_estimator: 200	1.0	0.95	0.96	0.96	0.96
Logistic Regression (TVEC)	C: 1.0	0.98	0.96	0.97	0.97	0.97
Adaboost (CVEC)	learning_rate: 1.0, n_estimator: 50	0.97	0.94	0.95	0.94	0.94
Adaboost (TVEC)	learning_rate: 1.0, n_estimator: 50	0.96	0.94	0.95	0.94	0.94
SVC (CVEC)	C: 1.0	0.99	0.94	0.78	0.66	0.53
SVC (TVEC)	C: 1.0	0.99	0.97	0.98	0.98	0.98

Error Analysis

		title_text	y_pred	y_test
1110	How to un-ignore notification Bit of an odd on...	applewatch	garmin	
1236	D2 Echo Anyone have any idea when this is comi...	applewatch	garmin	
1251	Virtual partner.. How do I do this? I have a ...	applewatch	garmin	
3274	Manually add workout to Activity app? I apolog...	garmin	applewatch	
3403	Best watch apps for working out at the gym? I'	garmin	applewatch	
2127	Constantly asks for permissions Had to delete ...	applewatch	garmin	
3389	Does it bother anyone else that the time isn't...	garmin	applewatch	
3508	Bug related to Longest Move Streak achievement...	garmin	applewatch	
879	How to get the g96 Mark Waypoint function to m...	applewatch	garmin	• • • •
2764	if your wifi network is hidden,the apple watch...	garmin	applewatch	• • • •
				• • • •
				• • • •

• • •
• • •
• • •
• • •

Error Analysis (Cont.)

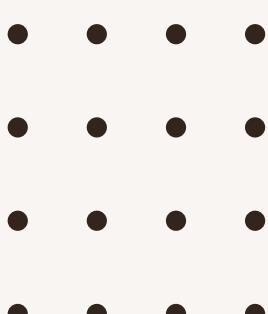
*“Just ordered a **935** after two many ruined **Apple watch** runs.
What features should I try out?...”*



Actual



Got Predicted



Conclusion

Best model is “Logistic Regression” with CVEC

The model has a train score of 0.98 and a test score of 0.96.

For confusion matrix:

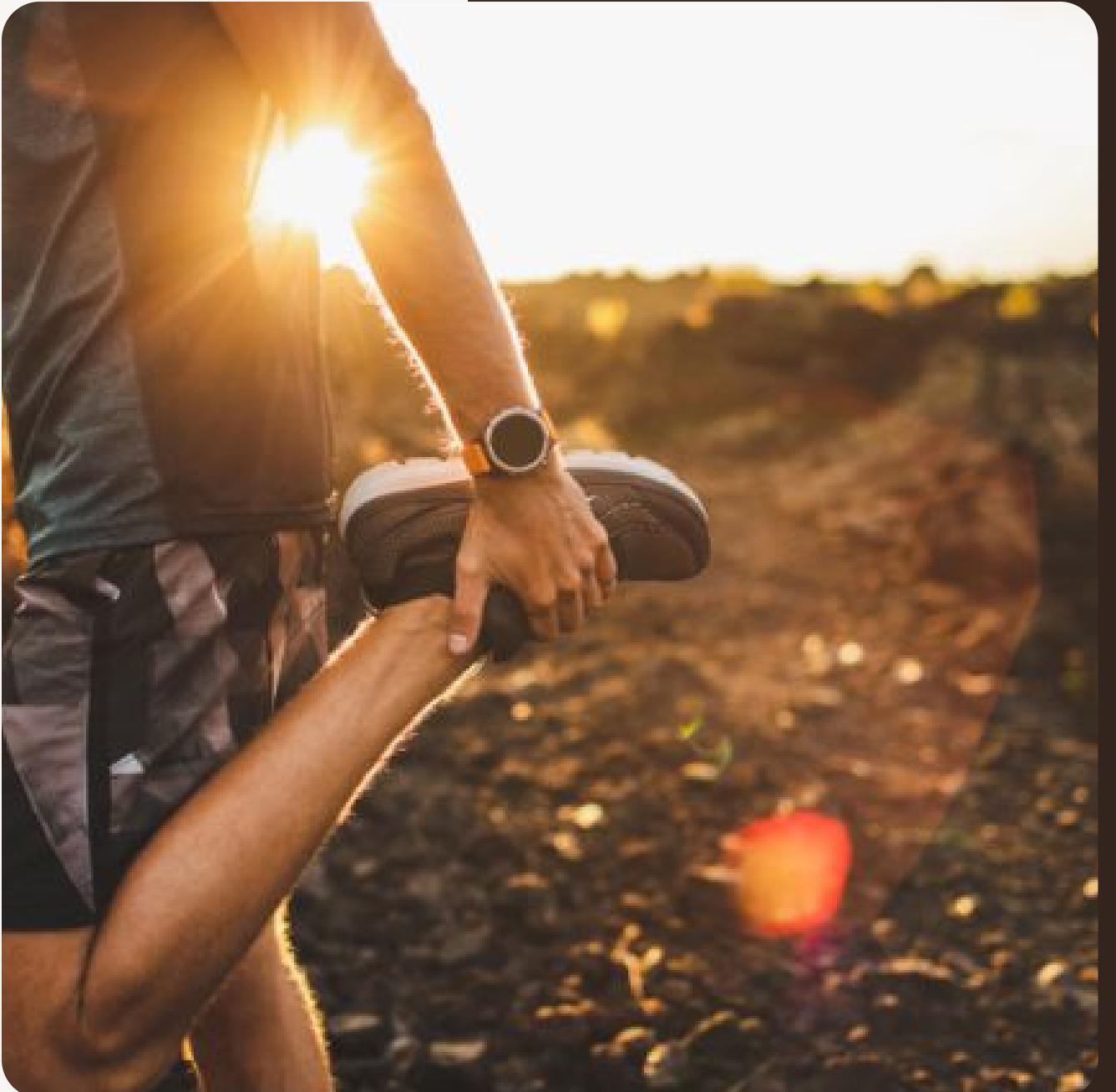
- precision score of 0.98
- recall score of 0.98.
- f1 score of 0.98.
- accuracy score of 0.98.

The model has a low variance of 0.01 and low bias of 0.01. The model is good fitting.



Recommendation

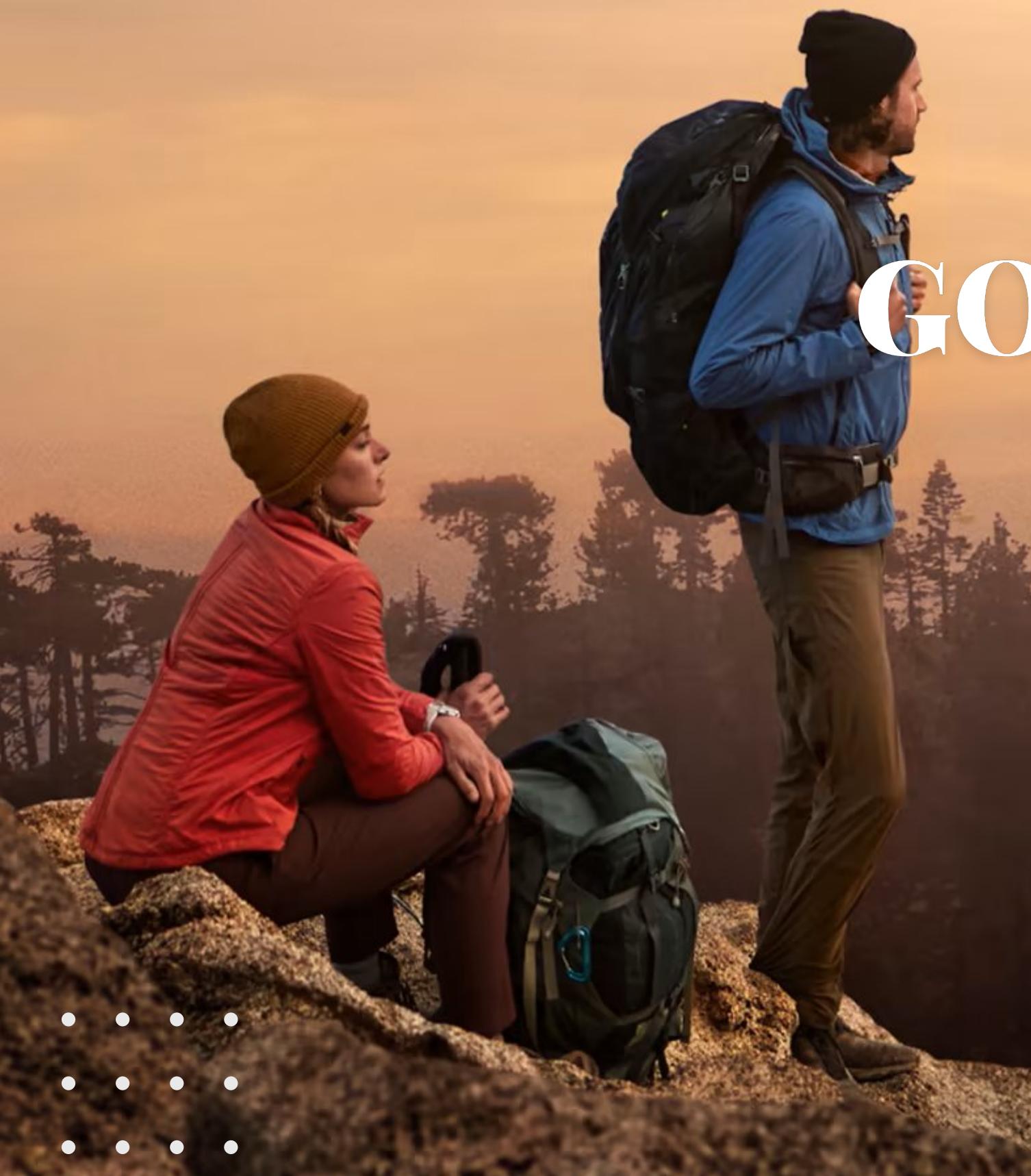
- To be make more prediction, tokenize and lemmatize can increase significant scores(precision, recall, F1 and other scores)
- Use Gaussian Naive Bayes, Voting, and other models to find other results, not only models that created.



Limitation

- When preprocessing on data from subreddit:
 - misspelled
 - slang words
 - misunderstandable words
- The context or the meaning of a sentence or phrase can depend on behavior of reddit users





COMFORT WHEN
GOING EVERYWHERE
ON A WATCH.

