Lexicon of Luxury-Vehicle Online Communities

Can trained models successfully classify language associated with vehicle affinity-groups?

Tesla Model X

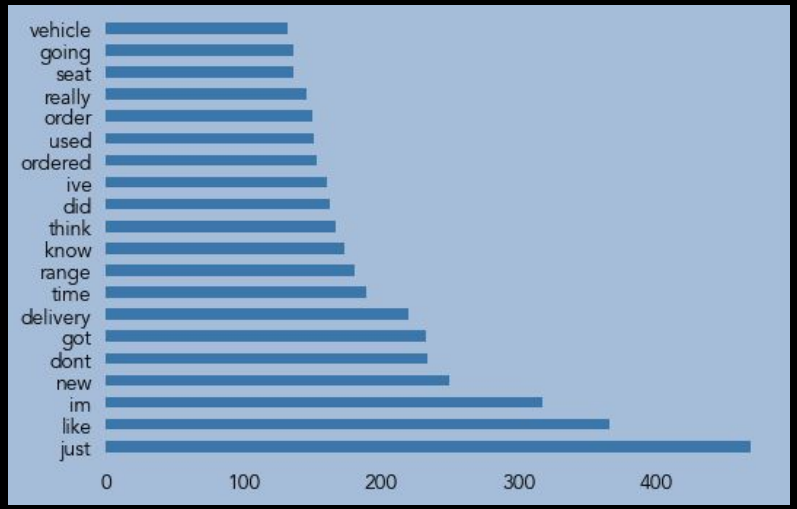
2015 | 2014 | 2.3k

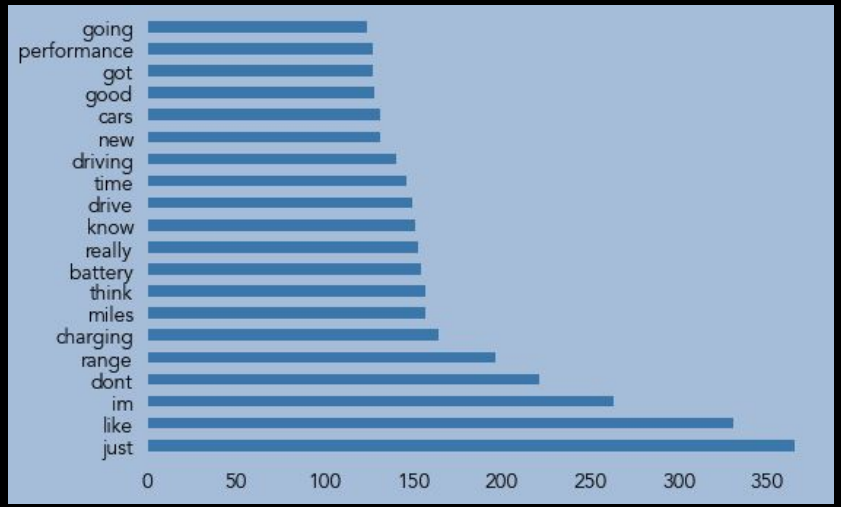
$ 115,000 - $138,000
avg: $126,000

Porsche Taycan, Turismo

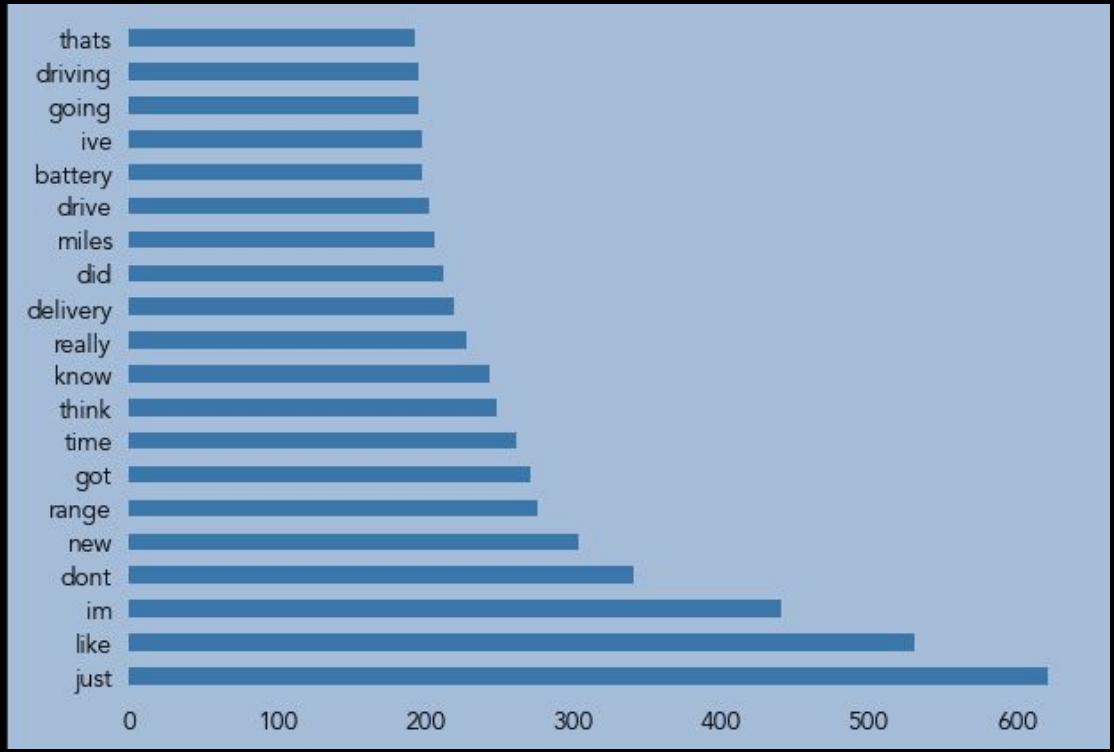
2019 | 2018 | 2.3k

$ 93,000 - $ 187,000
avg: $ 140,500


vehicle
going
seat
really
order
used
ordered
ive
did
think
know
range
time
delivery
got
dont
new
im
like
just
0
100
200
300
400

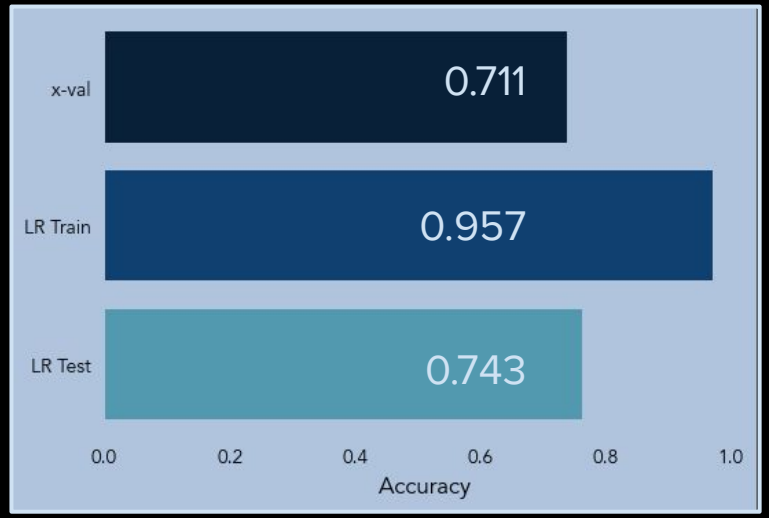

going
performance
got
good
cars
new
driving
time
drive
know
really
battery
think
miles
charging
range
dont
im
like
just
0
50
100
150
200
250
300
350

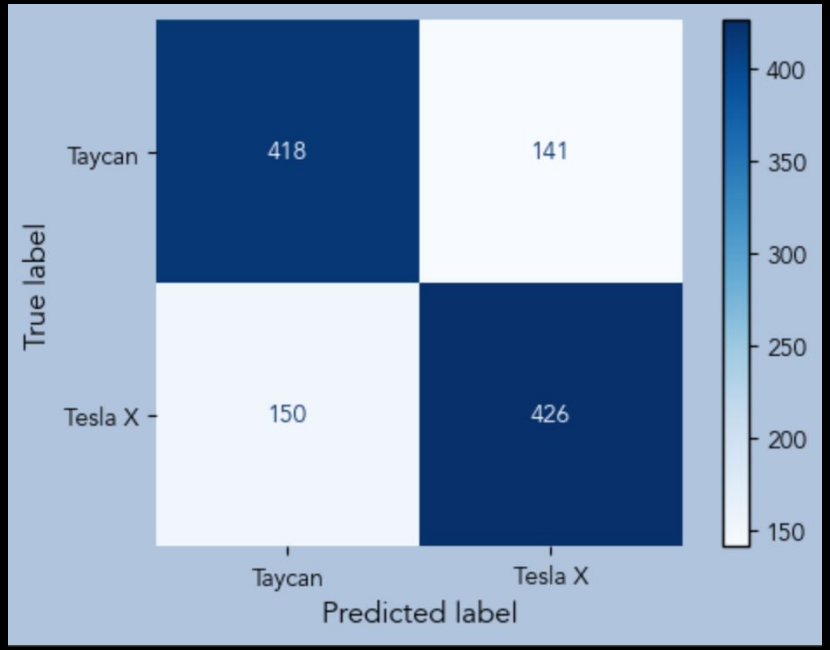
Word distribution, stop words



porsche
tesla
taycan
model
xplaid
turbo
turismo

Model 1 *n = 4 537*

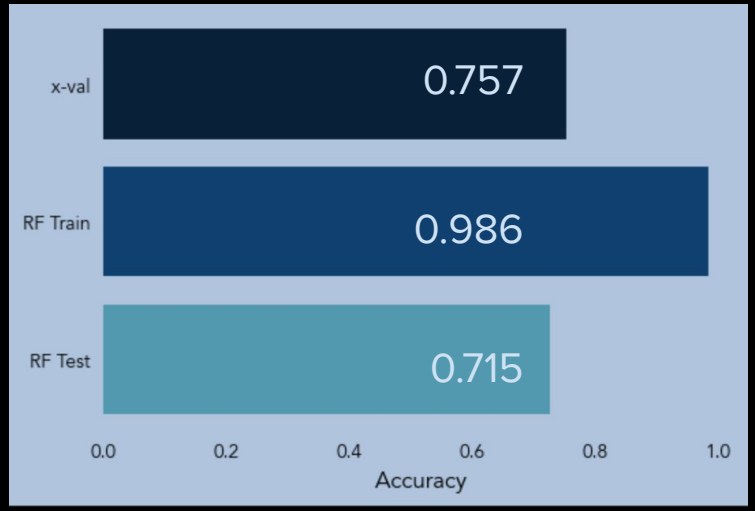


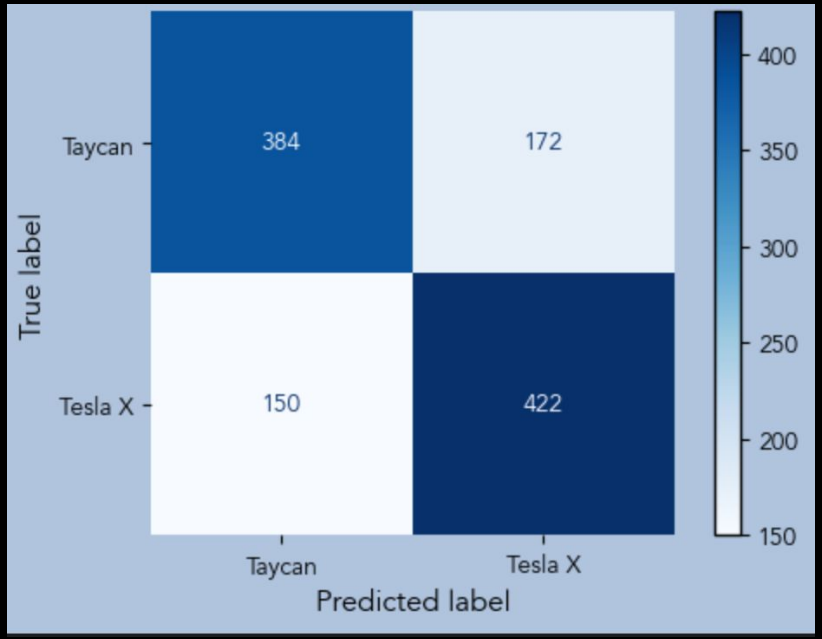


Transformer: Count Vectorizer
Estimator: Logistic Regression
Parameters: Default

- *mine just moved to july this morning*
- *this looks amazing*
- *that is confirmed to be the case*
- *this couldn't agree more*

Model 2 *n = 4 537*

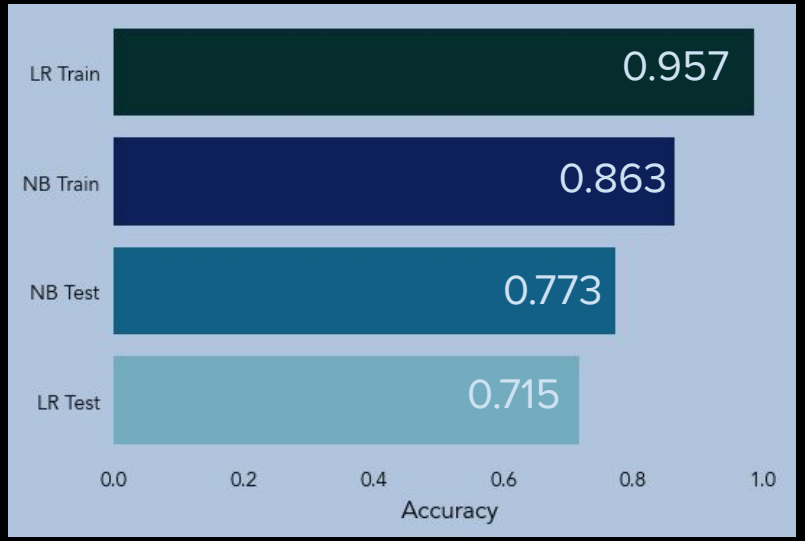




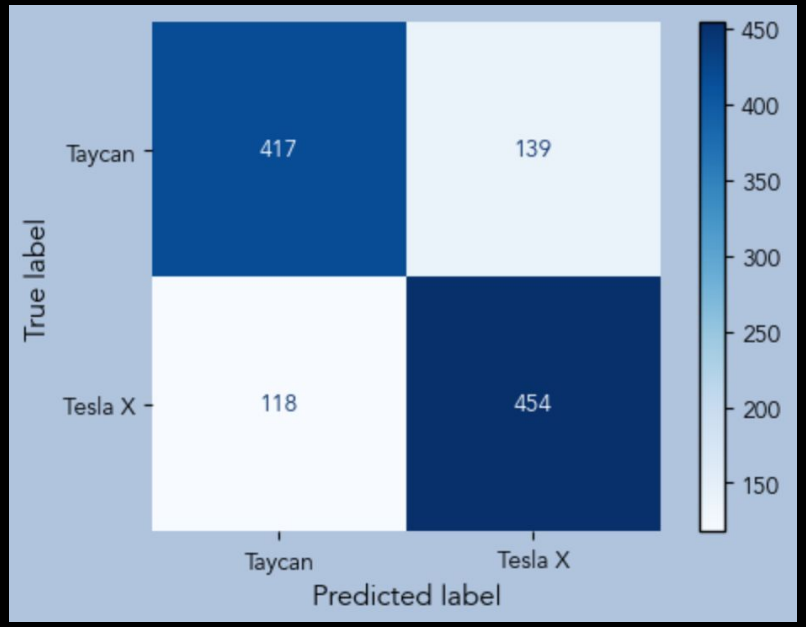
Transformer: Count Vectorizer
Estimator: Random Forests
Parameters: Default

i think about it some of my favorite conversations with my daughter when were cruising on the hwy and shes just making random observations about what she sees out the window i would be a fool to cede that to peppa pig voluntarily

Model 3 *n = 4 537*

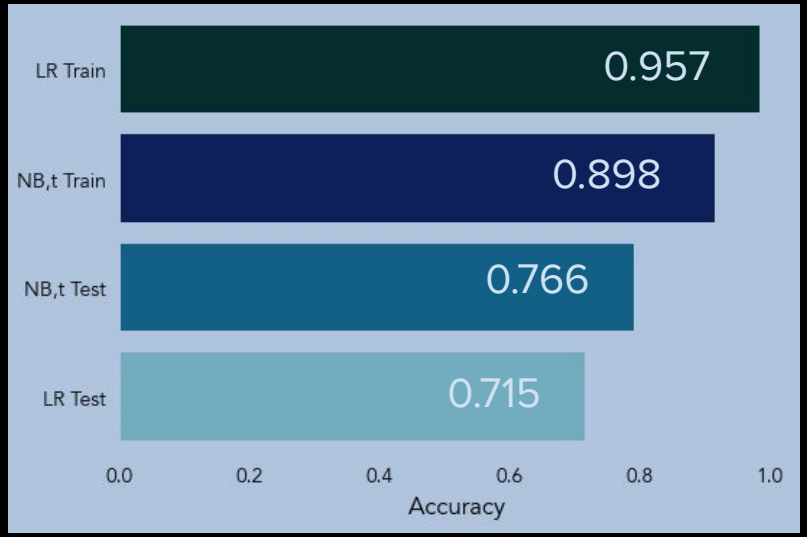


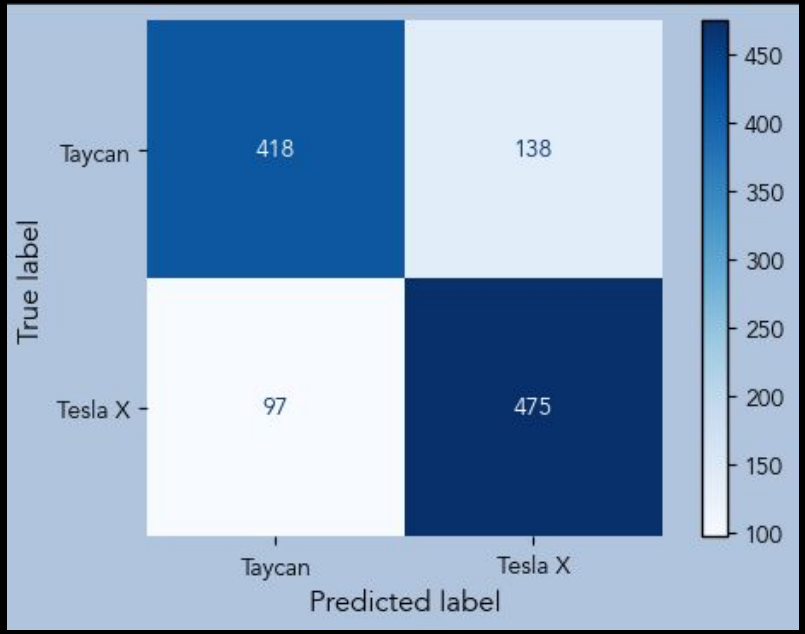
Transformer: Count Vectorizer
Estimator: Naïve Bayes
Parameters: Default



“i think its debatable as to whether its technically allowed under current laws but i think it will be terrible in practice really”

Model 4 *n = 4 537*

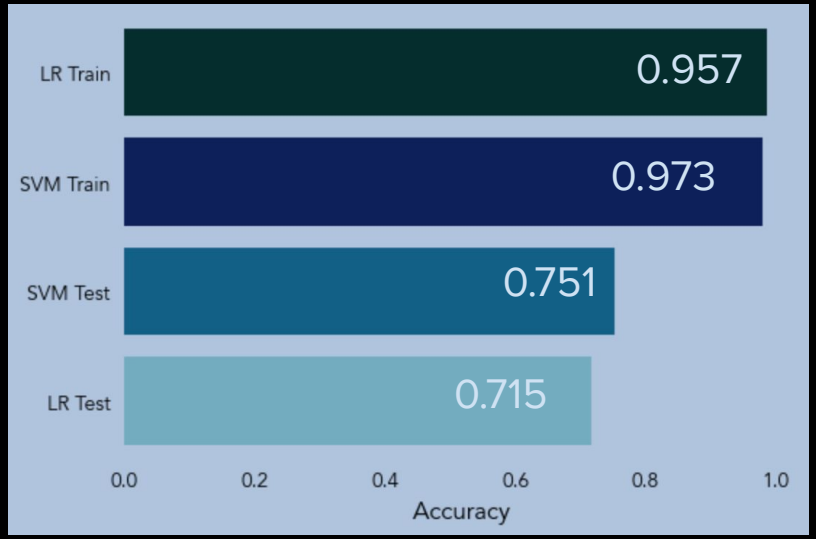


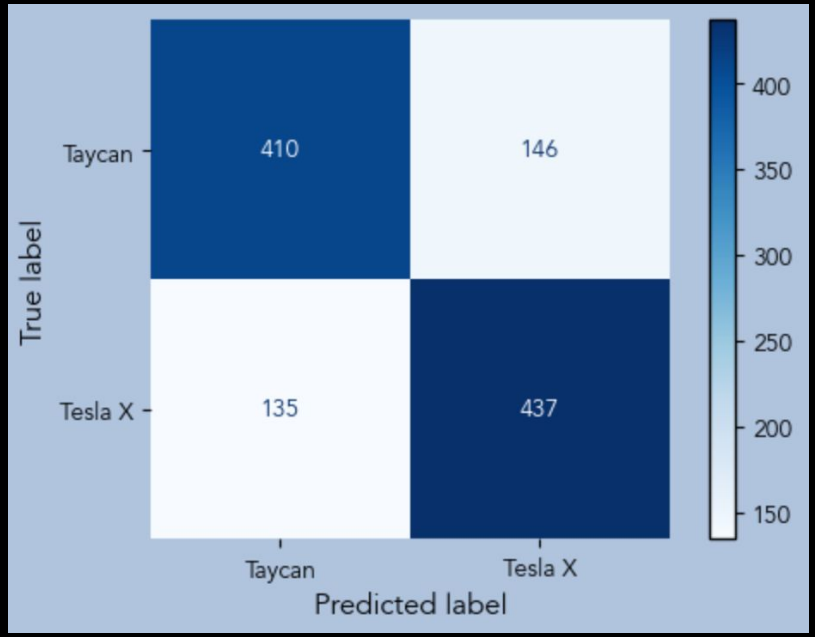


Transformer: TF-IDF
Estimator: Naïve Bayes
Parameters: Default

"i think its debatable as to whether its technically allowed under current laws but i think it will be terrible in practice really"

Model 5 *n = 4 537*

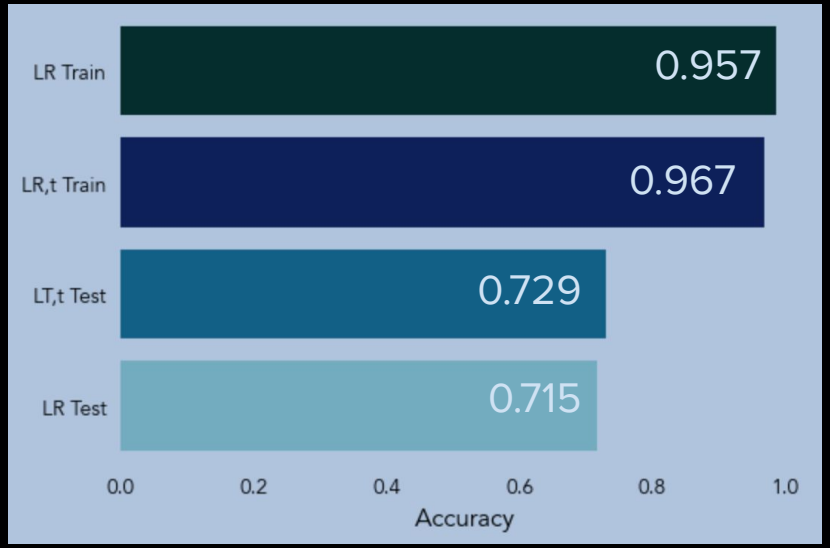




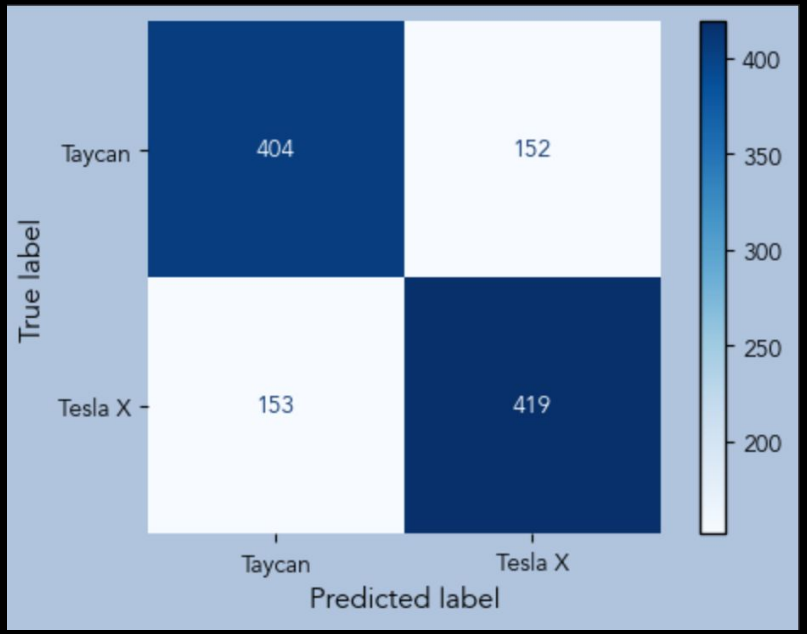
Transformer: TF-IDF
Estimator: SVM
Parameters: Default

wtf this subreddit is for the ***tesla model*** *x what is this*

Model 6 *n = 4 537*

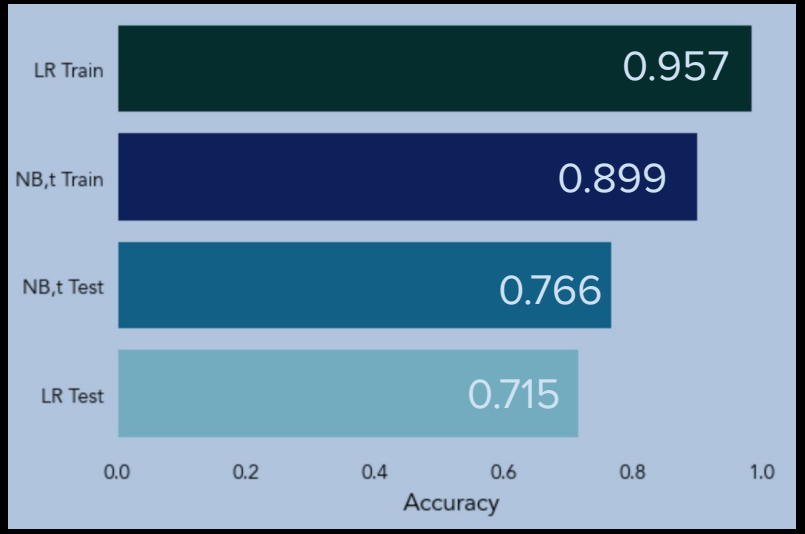


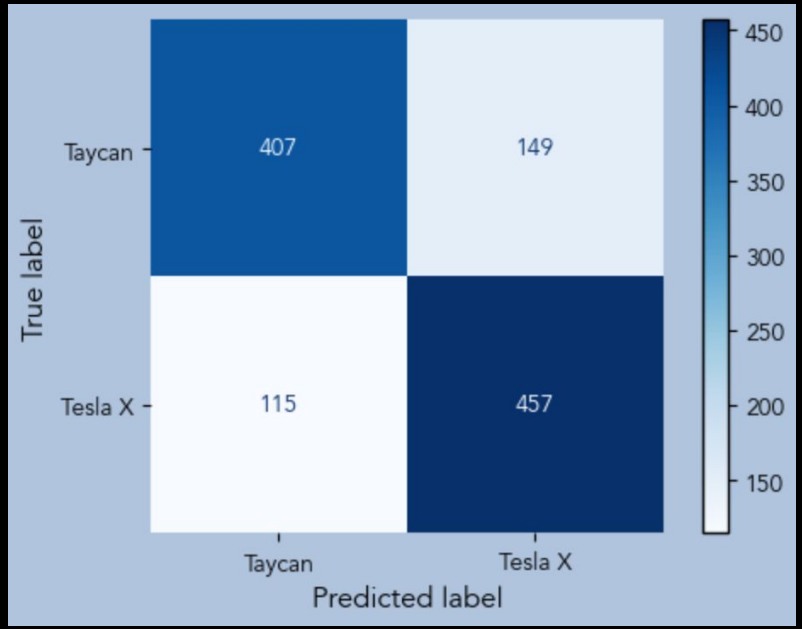
Transformer: TF-IDF
Estimator: Logistic Regressor
Parameters: Default



i dont know all the specifics except that they have to update multiple computers and that sometimes it doesnt take and they have to do a complete reset rebuild

Model 7 *n = 4 537*

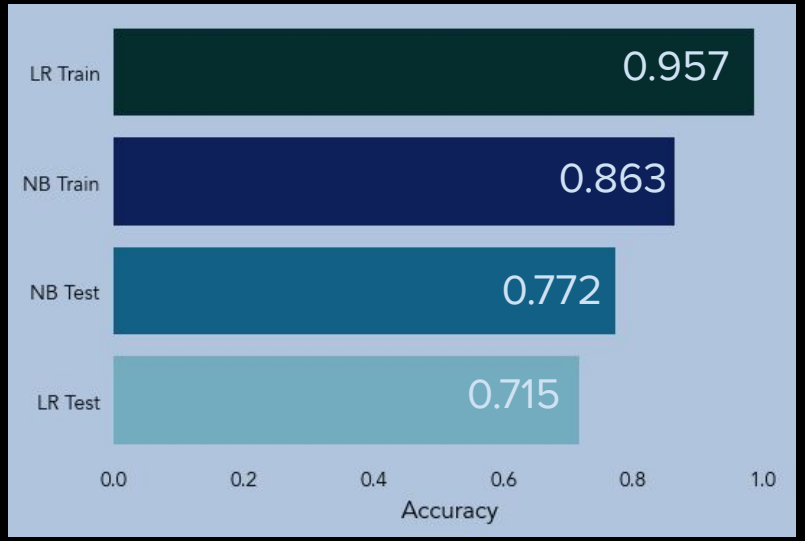


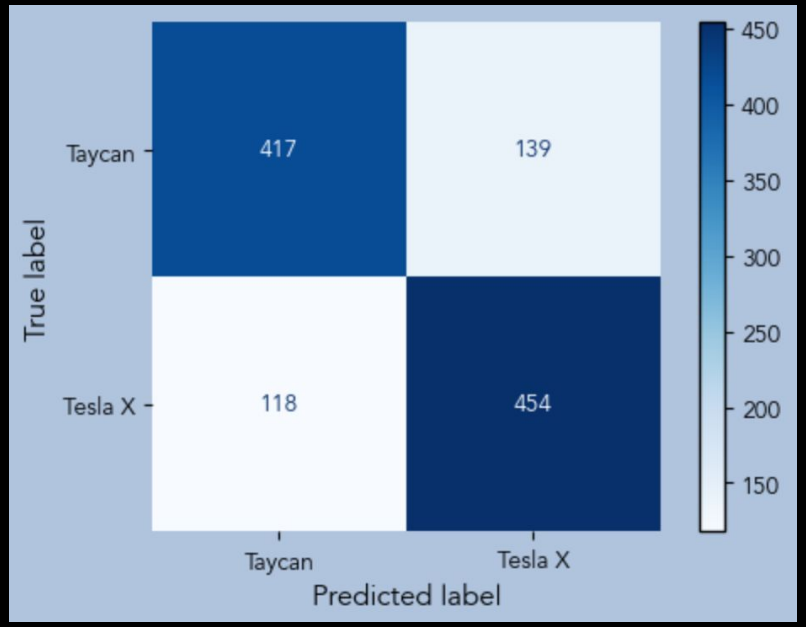


Transformer: TF-IDF
Estimator: Naïve Bayes
Parameters: Default

not according to my research do you have anything to back your claim up

Model 8 *elon*





Transformer: Count Vectorizer
Estimator: Naïve Bayes
Parameters: Default

i think they should at least be less conservative

Parameters

max features	2000, 3000, 4000, ***7000***
min df	***2***, 3, 4
max df	***0.7***, 0.8, 0.9, 0.95
n_gram	(1,1), (1,2),**(1,3)**

Conclusion

- Language distinct enough for further analysis
- Source data from different venues