Project 3 :

NLP - Mercedes-Benz and BMW

Nov 13, 2023 : DSI 03

Problem Statement

“ A car trading agency company in England having two main premium-class cars which are Mercedes-Benz and BMW. In a day, the customer service team have inbox emails over a thousand from senders throughout the U.K. During the expansion of branches in this year, the number of the inbox email is increasing, so the head of customer service asks Data Science team to create model in order to distinguish the email by vehicle brands and send to the incharing section directly. As company-email gathering seems confidential, so the data science team started to do web-scraping from Reddit to initiate model creation instead at first. ”

Data Overview

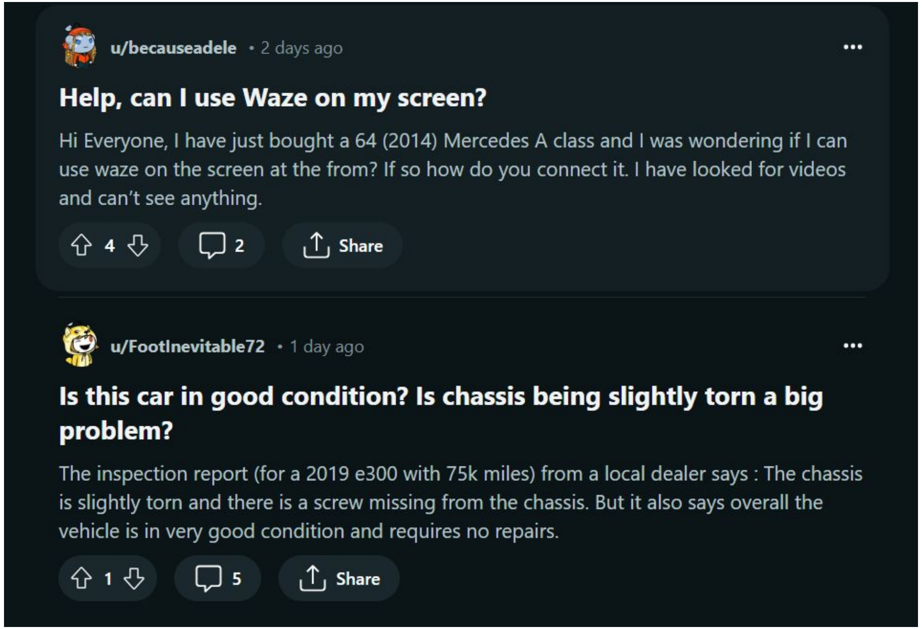
- 24,075 observations from Mercedes-Benz sub-reddit, gathered from 2012 to 2023, and 31,435 observations from BMW sub-reddit, gathered from 2010 to 2018.

- The observations which contains no post contents, having only title, were dropped.

- After considering the above condition, the observations on dataset remains 13,030 (48% as Mercedes-Benz and 52% as BMW)

- `‘post’` columns are added by concating `‘title’` and `‘text columns’` columns

Post illustrations



u/becauseadele • 2 days ago

Help, can I use Waze on my screen?

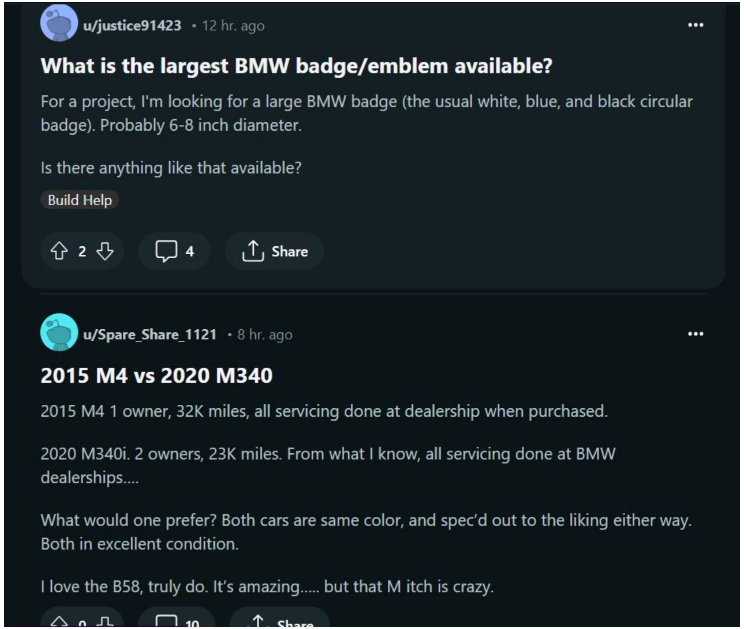
Hi Everyone, I have just bought a 64 (2014) Mercedes A class and I was wondering if I can use waze on the screen at the from? If so how do you connect it. I have looked for videos and can't see anything.

u/FootInevitable72 • 1 day ago

Is this car in good condition? Is chassis being slightly torn a big problem?

The inspection report (for a 2019 e300 with 75k miles) from a local dealer says : The chassis is slightly torn and there is a screw missing from the chassis. But it also says overall the vehicle is in very good condition and requires no repairs.

Posts from Mercedes-Benz sub-reddit



u/justice91423 • 12 hr. ago

What is the largest BMW badge/emblem available?

For a project, I'm looking for a large BMW badge (the usual white, blue, and black circular badge). Probably 6-8 inch diameter.

Is there anything like that available?

u/Spare_Share_1121 • 8 hr. ago

2015 M4 vs 2020 M340

2015 M4 1 owner, 32K miles, all servicing done at dealership when purchased.

2020 M340i. 2 owners, 23K miles. From what I know, all servicing done at BMW dealerships....

What would one prefer? Both cars are same color, and spec'd out to the liking either way. Both in excellent condition.

I love the B58, truly do. It's amazing..... but that M itch is crazy.

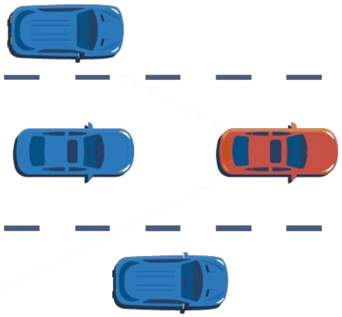
Posts from BMW sub-reddit

Model Pre-processing

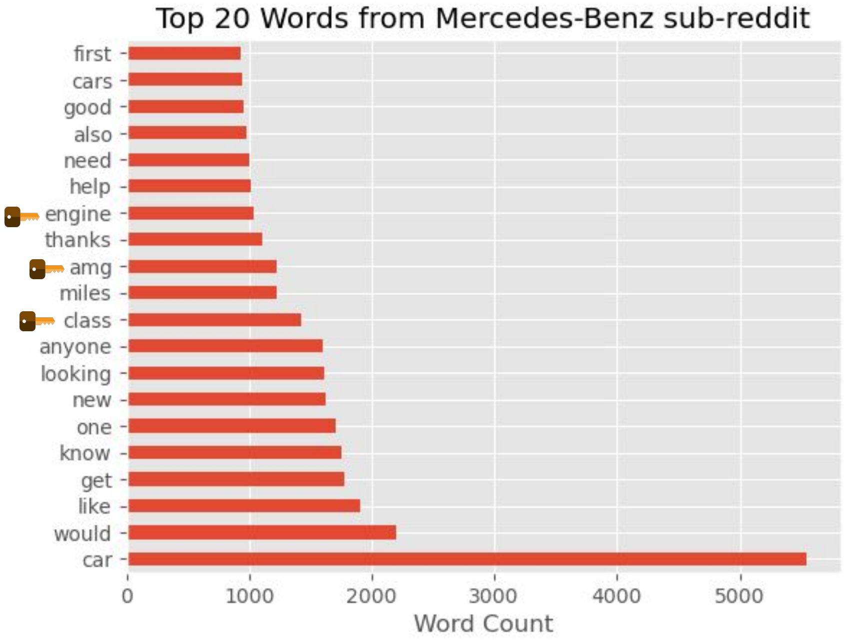
- Remove duplicate observations

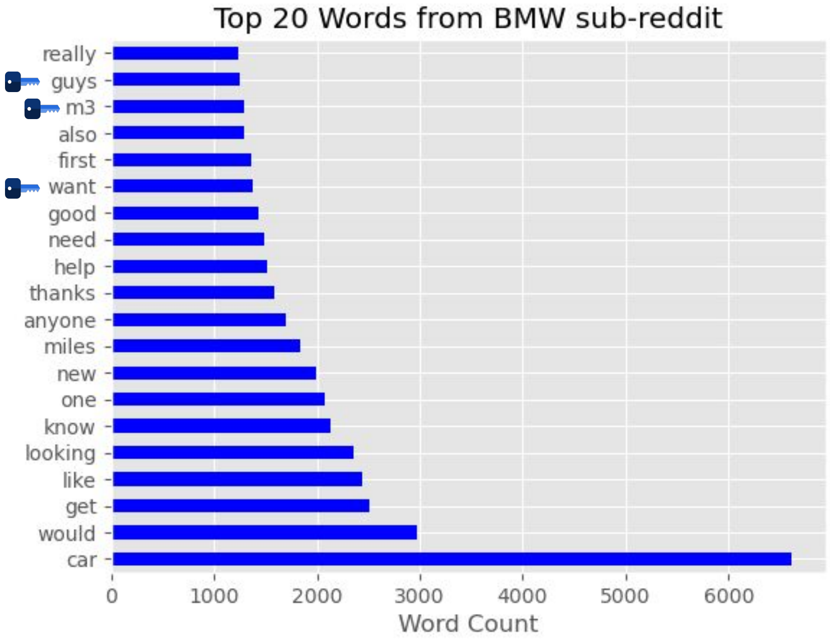
- `['BMW','Mercedes-Benz','Benz','Mercedes','MB']` are added to English stop words which the models ignore in the classification.

- `CountVectorizer` are applied to the models



`‘class’` is one of most common words for Mercedes-Benz and `‘m3’` is one of most common words for BMW


Top 20 Words from Mercedes-Benz sub-reddit


Top 20 Words from BMW sub-reddit

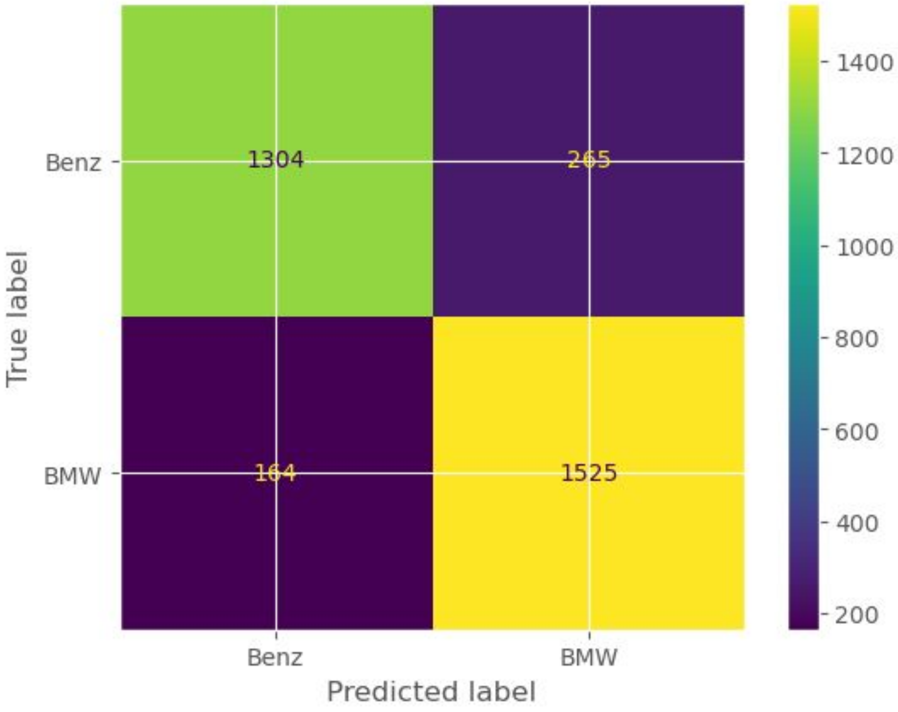
Model 1 : Naive Bayes - Performance

- Accuracy score on Training : 0.87
- Accuracy score on Test : 0.87

- Recall (Mercedes-Benz) : 0.84
- Recall (BMW) : 0.90

- Precision (Mercedes-Benz) : 0.89
- Precision (BMW) : 0.85

- F-1 Score (Mercedes-Benz) : 0.86
- F-1 Score (BMW) : 0.88



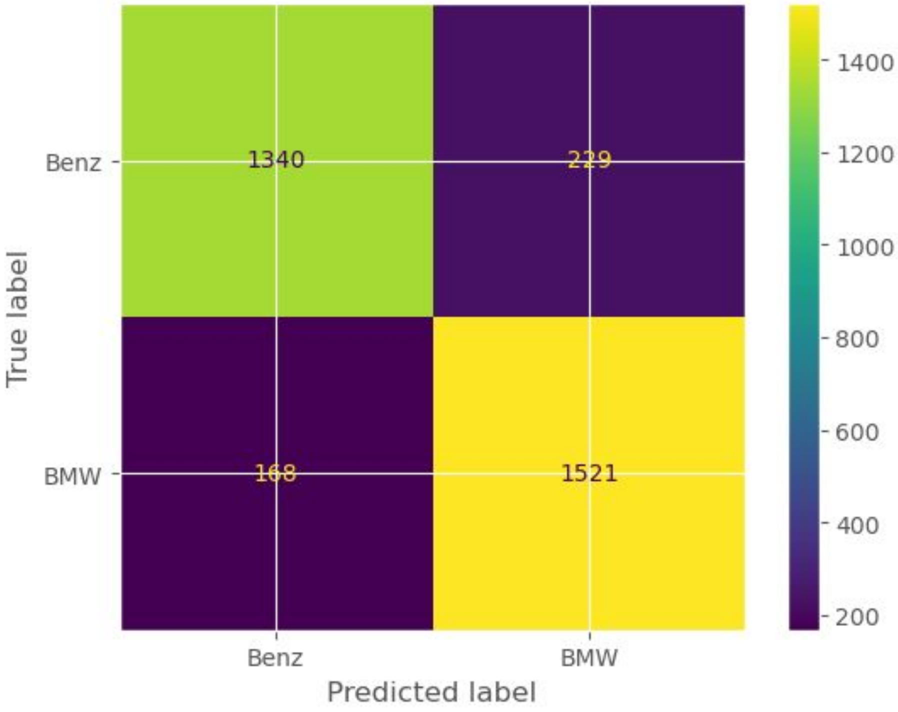
Model 2 : Random Forest with GridSearchCV - Performance

- Accuracy score on Training : 0.99
- Accuracy score on Test : 0.88

- Recall (Mercedes-Benz) : 0.87
- Recall (BMW) : 0.89

- Precision (Mercedes-Benz) : 0.88
- Precision (BMW) : 0.88

- F-1 Score (Mercedes-Benz) : 0.87
- F-1 Score (BMW) : 0.88

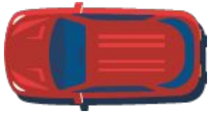


Error Visualizing

Actual class is ‘Mercedes-Benz’ but predicted as BMW	Actual class is ‘BMW’ but predicted as Mercedes-Benz
Is a running W140 with 200k miles for $800 a good buy? I have the opportunity to buy a running w140 with 200k miles for $800. I’m not very mechanically inclined but I'm willing to learn. Do you **guys** think it will be worth it as a second/project car?	Thoughts on the 04 525i I'm looking at purchasing a 525i, less than 100k on the odo, looks good. From a dealership used, 1 owner. Any habitual problems with this **engine**, or drive train that reddit knows of? The reviews online are all garbage (fanboys or evil experiences.) Hoping for some good feedback.
What car wash supplies do you recommend? I now have the time to wash and detail my car myself every couple weekends. I was wondering what wash, wax, towels etc you **guys** recommend. My car is White	Went to First **Class** Fitment (Canibeat) I saw some great looking BMW's, I thought I should share! Sorry for the quality, was taken from my cellphone. Gallery

Model Comparing

	Model 1 : Naive Bayes	**Model 2 : Random Forest with GridSearchCV**
Accuracy on Training	0.87	0.99
Accuracy on Test	0.87	0.88
Recall (Mercedes-Benz)	0.84	0.87
Recall (BMW)	0.90	0.89
Precision	0.87	0.88
F-1 Score	0.87	0.88



Conclusion and Recommendation

The Data Science team recommended Model 2 : Random Forest with GridSearchCV to customer service team which has an accuracy 88% on unseen data set. If the head of customer service accept with this model, the customer service team need to verify for the other 12% of the incoming emails.