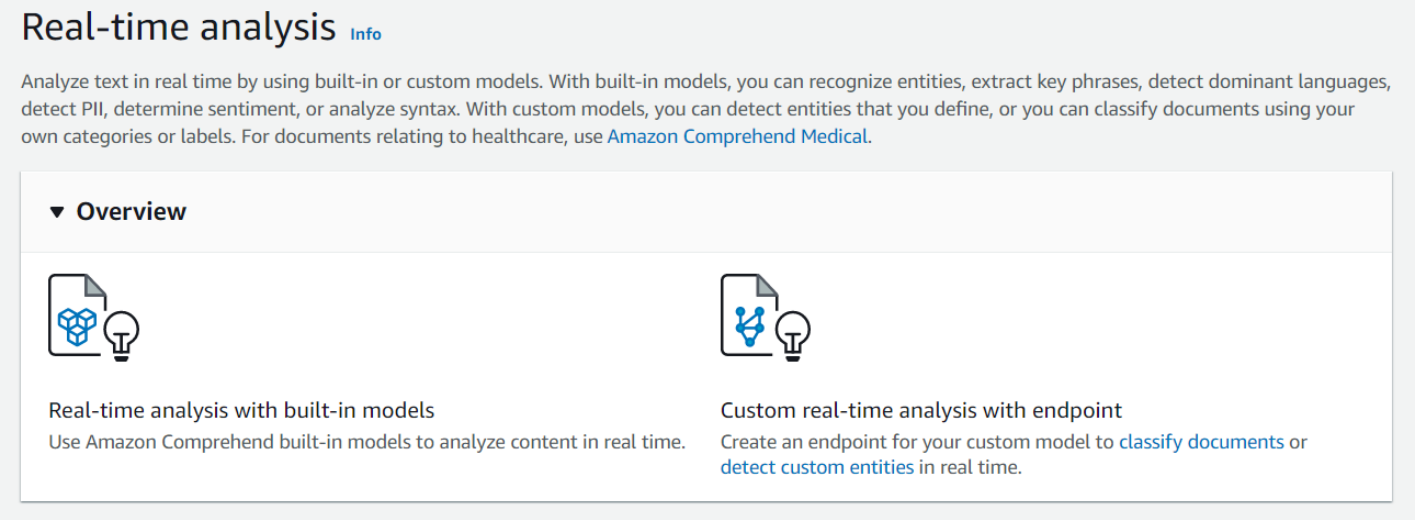
PROJECT - 5

Uploaded CSV file for optional task

PART – I

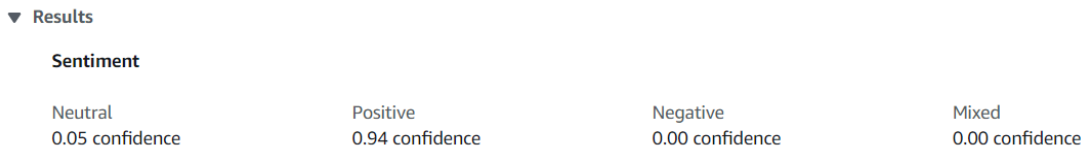
Q1) You are given the extracted csv having 20 reviews of the above product. Analyze the first 10 reviews independently using Amazon Comprehend and answer the below questions.



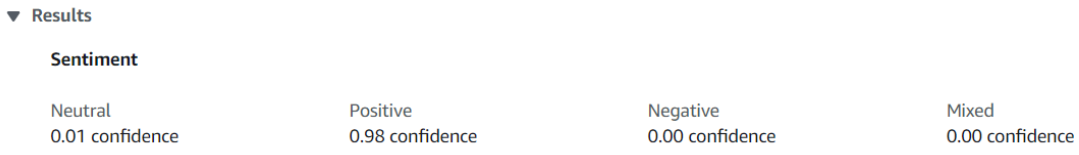
Performing built-in analysis on each review

a. Report the sentiment and its confidence percentage for each comment. (5 points)

1.



2.



3.

4.

5.

Sentiment

Neutral	Positive	Negative	Mixed
0.00 confidence	0.99 confidence	0.00 confidence	0.00 confidence

6.

Sentiment

Neutral	Positive	Negative	Mixed
0.00 confidence	0.69 confidence	0.00 confidence	0.30 confidence

7.

Sentiment

Neutral	Positive	Negative	Mixed
0.00 confidence	0.00 confidence	0.06 confidence	0.93 confidence

8.

Sentiment

Neutral	Positive	Negative	Mixed
0.00 confidence	0.99 confidence	0.00 confidence	0.00 confidence

9.

Sentiment

Neutral	Positive	Negative	Mixed
0.00 confidence	0.99 confidence	0.00 confidence	0.00 confidence

10.

Sentiment

Neutral	Positive	Negative	Mixed
0.05 confidence	0.93 confidence	0.00 confidence	0.00 confidence

11.

Sentiment

Neutral	Positive	Negative	Mixed
0.00 confidence	0.99 confidence	0.00 confidence	0.00 confidence

12.

Sentiment

Neutral	Positive	Negative	Mixed
0.00 confidence	0.01 confidence	0.00 confidence	0.98 confidence

13.

Sentiment

Neutral	Positive	Negative	Mixed
0.00 confidence	0.99 confidence	0.00 confidence	0.00 confidence

14.

Sentiment

Neutral	Positive	Negative	Mixed
0.00 confidence	0.11 confidence	0.00 confidence	0.87 confidence

15.

Sentiment

Neutral	Positive	Negative	Mixed
0.00 confidence	0.22 confidence	0.02 confidence	0.75 confidence

16.

Sentiment

Neutral	Positive	Negative	Mixed
0.28 confidence	0.08 confidence	0.62 confidence	0.00 confidence

17.

Sentiment

Neutral	Positive	Negative	Mixed
0.00 confidence	0.87 confidence	0.00 confidence	0.12 confidence

18.

Sentiment

Neutral	Positive	Negative	Mixed
0.05 confidence	0.00 confidence	0.92 confidence	0.00 confidence

19.

Sentiment

Neutral	Positive	Negative	Mixed
0.00 confidence	0.98 confidence	0.00 confidence	0.00 confidence

20.

Sentiment

Neutral	Positive	Negative	Mixed
0.00 confidence	0.99 confidence	0.00 confidence	0.00 confidence

b. Report the overall summary of the 10 comments. The summary should contain the percentage of negative, positive, mixed, and neutral comments. (2 points)

Ans.

Sentiment

Neutral	Positive	Negative	Mixed
0.05 confidence	0.94 confidence	0.00 confidence	0.00 confidence

last last google home	0.99+
India	0.99+
The Launch Day	0.99+
the Free JioFi	0.99+
Proper Packing	0.88
Next Day Delivery	0.97
a pretty good Virtual Assistant	0.99+
many useful functions	0.99+
the useful ones	0.99+
an Alarm	0.99+

<- Key phrase

SUMMARY GENERATOR

Free online text summarizer based on open source text summarization software.

Amazing assistant by Google ... Have many features like, using Wikipedia by voice command, getting information on any topic, playing any songs as and when u requested, and so many features like quiz on various topics, learning English and traffic at ur home and work location..... And google Still adding many more features in Google Home App..... Works with power bank too.... Great Product ...Is still not 100% perfect but still is a great buy.

Summarize Copy Clear

Here is also summary generated online tool available.

Sentiment

Neutral	Positive	Negative	Mixed
0.00 confidence	0.99 confidence	0.00 confidence	0.00 confidence

There is also a different version of summary generated and here are its sentiment ratios

"At last last google home is in India.

You will be able to do many things with your Google Home or Echo devices then.

I have both Google Home Mini and Amazon Echo Dot.

Amazing assistant by Google ... Have many features like, using Wikipedia by voice command, getting information on any topic, playing any songs as and when u requested, and so many features like quiz on various topics, learning English and traffic at ur home and work location..... And google Still adding many more features in Google Home App..... Works with power bank too.... Great Product ...Is still not 100% perfect but still is a great buy.

It doesn't always recognize's the lyrics of the song with English (United States) but works perfectly fine with English (India) "Google home mini is very good device and a smart speaker.

Problem arises with its mic itself we can't control home mini when it plays a song command '"hey Google"' it just can't recognise you need to scream louder and It then recognises sometimes not even when i scream .

Google home in India lacks lot of featu..." This is an awesome device.

Sentiment

Neutral	Positive	Negative	Mixed
0.02 confidence	0.77 confidence	0.00 confidence	0.19 confidence

Overall it is a positive text as we have really less percentage of negative comments.

c. Comment whether the majority of comments are positive or negative? (1 points)
Ans.
Yes, majority of comments are positive as we can see both from individual analysis and summary analysis. We can notice that either the comment in top 10 comments is positive or is giving a pros and cons type. That is have mixed emotions. There is no completely negative comment at all present.

Q2) Analyze all the negative & mixed comments from the scraped reviews and answer the below questions.

I am analyzing comments/reviews 4, 7, 12, 14, 15 and 16. I have considered the comments have more then 50% of negative or mixed confidence percentages for my basis of dividing and selecting them for analysis so that I can get the correct customer sentiment analysis.

a. What are the common key phrases in these comments? Report at least 10 key words. (1 points)
Ans.

Key phrases	Confidence
Chromecast	0.99+
youtube videos	0.99+
external speaker	0.99+
calls	0.99+
Google	0.99+
many features	0.99+
Wikipedia	0.99+
voice command	0.99+
information	0.99+
any topic	0.99+

For negative comments:

Key phrases	Confidence
No 3.5 mm	0.99+
external speaker	0.99+
calls	0.99+
USA , Canada	0.99+
India	0.99+
Problem	0.99+
its mic	0.99+
home mini	0.99+
a song command	0.99+
hey Google	0.85

b. Report part of speech of this first 10 words of key phrases of the previous comment. (1 points)

Ans.

1 and 2 - No 3.5 mm jack to connect to external speaker.

3 - Not able to make calls.

4 - it works trumendouly well in USA , Canada But in India it's kinda disappointing. Problem arises with its mic itself we can't control home mini when it plays a song command ""hey Google"" it just can't recognise you need to scream louder and It then recognises sometimes not even when i scream.

5 - Honestly, this is just an interactive speaker for now in India.

6, 7, 9 and 10 - Problem arises with its mic itself we can't control home mini when it plays a song command ""hey Google"" it just can't recognise you need to scream louder and It then recognises sometimes not even when i scream. It can also understand Hindi now, although not very well yet.

8 - Maybe the regular Google Home does it better but from my understanding, both have similar hardware for picking up commands.

c. List down the common problems or areas of improvement in the product, by analyzing the key phrases of negative comments. (3 points)

Ans.

- India version of Google Home Mini has not a good voice recognition and noise reduction.
- There is no 3.5 mm Jack present on the device.
- There is not option for plugin we have to either use battery or power bank. Need to have power source feature.
- Can't recognize US English songs when set to Indian English or Hindi songs when language not set (Should be able to recognize and analyze the language accordingly (language translation needed to be integrated))
- Not able to recognize some Indian brands and commonly used words.
- Can't update firmware in our own in case of updates.
- No return policy from Flipkart, just replacement available.
- Not able to connect Bluetooth speaker to Google Mini.
- Need to update Indian Content.

d. List down suggestions to improve the product which will result in better customer satisfaction. (2 points)

Ans.

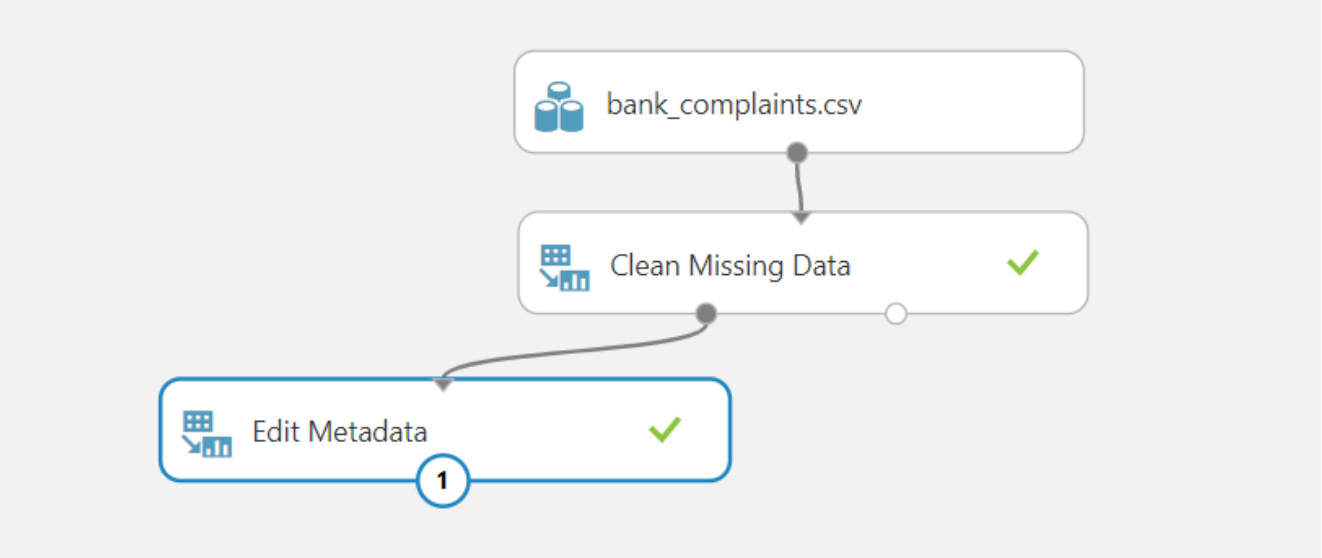
- Adding the feature of power source plugin (can be used without battery or power bank) and 3.5 mm jack or Bluetooth speaker connectivity to enhance the audio experience for the customer.
- Making return policy available or displaying pop up notices for region specifics features are at times not available for customers of certain region.
- Updating Google Mini according to region needs (adding region content as mentioned in case of India it's still not content having all features nor is updated with Indian content).
- Should be able to recognize and analyze the language accordingly (language translation needed to be integrated))
- Making firmware updates easily accessible and available for customers.

PART – II

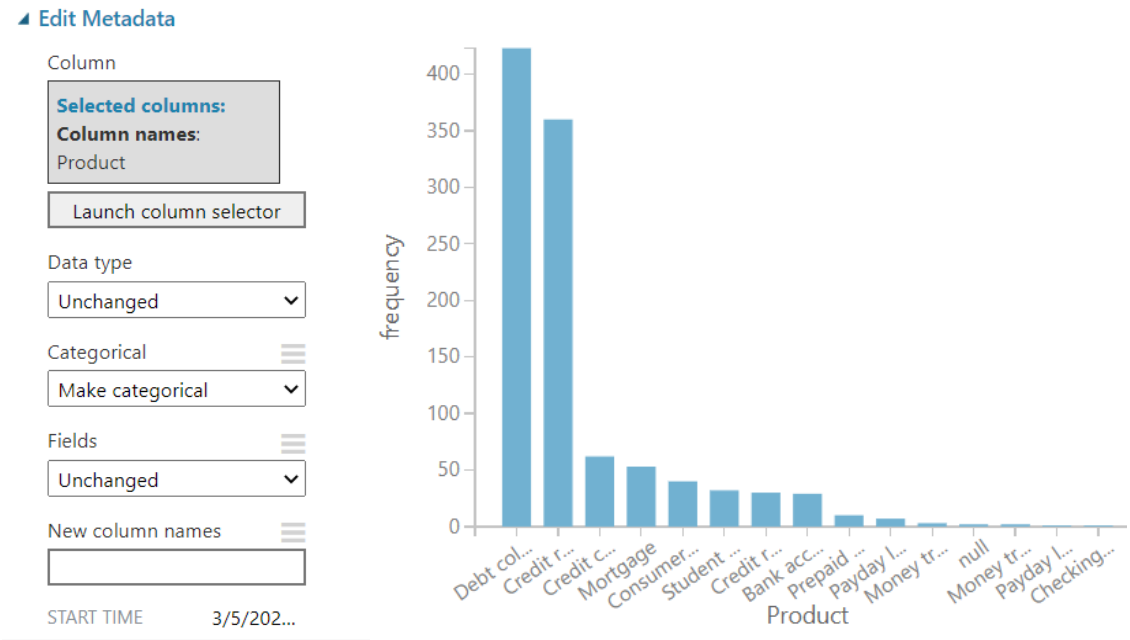
Complaint Categorization Baseline Model

Q1) Upload the provided dataset in AZURE ML and convert the "product" feature into categorical data.

Ans.

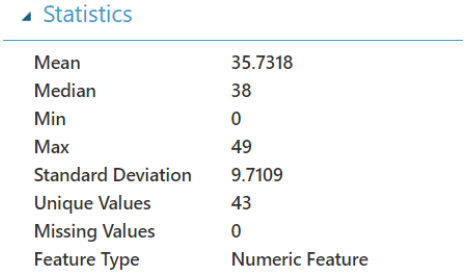


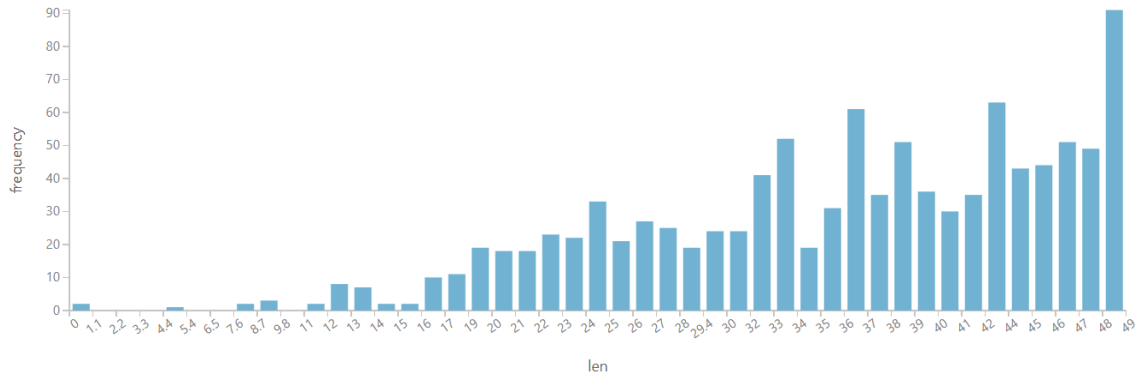
Made product Categorical after cleaning missing values



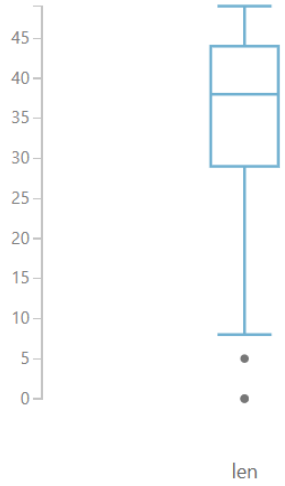
Q2) Visualize the “Len” column and report min and max values of the length. Comment on the distribution of the “Len” feature. (1 points)

Ans.





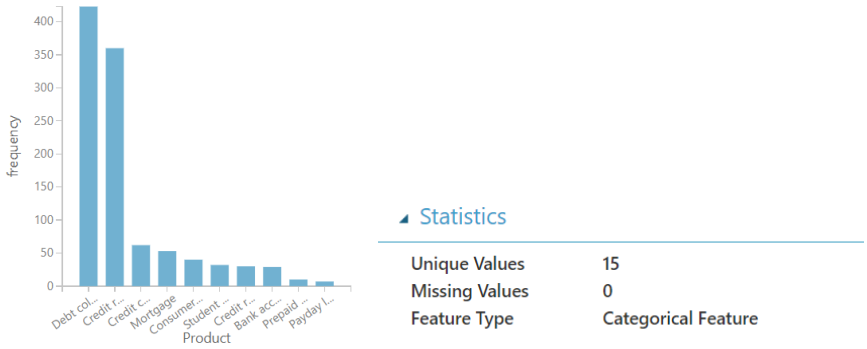
Since box plot helps in better analyzing of data, we can see that mean is 35.7318, median 38 and min and max values being 0 and 49 respectively.



We can also observe that most of the lengths are above 25 and there are really few lengths below 10 with 2 values of Len being exceptionally low which might be an invalid comment.

Q3) How many unique products are there? Which product has the highest number of complaints? (1 points)

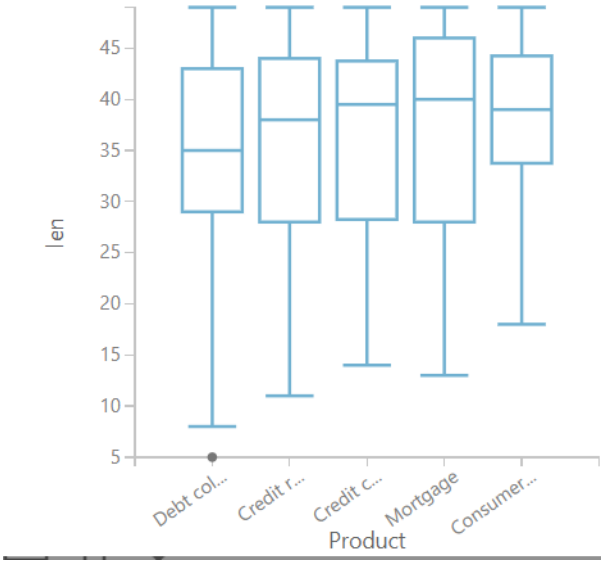
Ans.



There are 15 unique products and Debt collection has highest number of Complaints.

Q4) It is said that the average length of comments related to the “debt collection” product is comparatively lower than other products. Is this statement true? Provide necessary evidence to prove/disprove this claim. (2 points)

Ans. Yes, this is true that the average length of comments related to debt collection product is lower compared to other products.



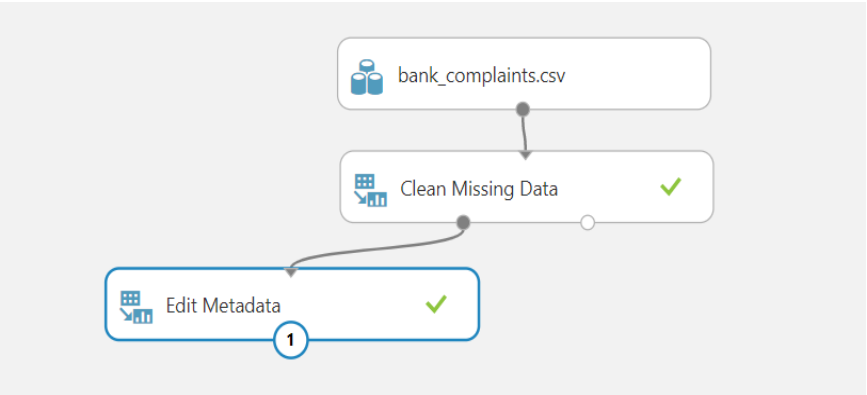
As we can see from this box plot the mean or the average line is lower then that of others and also there is exceptionally low value of length also existing as indicated by the dot.

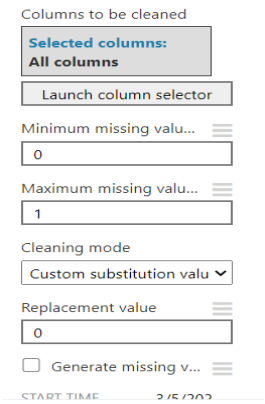
This hypothesis is true.

Q5) Check for null values and drop them if there are any. (1 point)

Ans.

I have already checked the null values and dropped them in the first step itself. There were 2 null values in both product and Len columns.

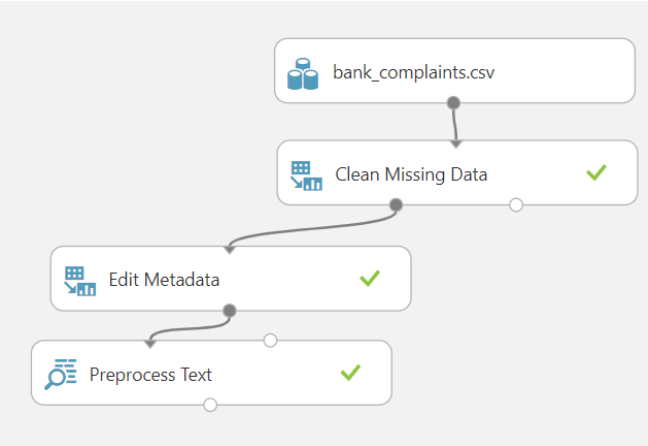


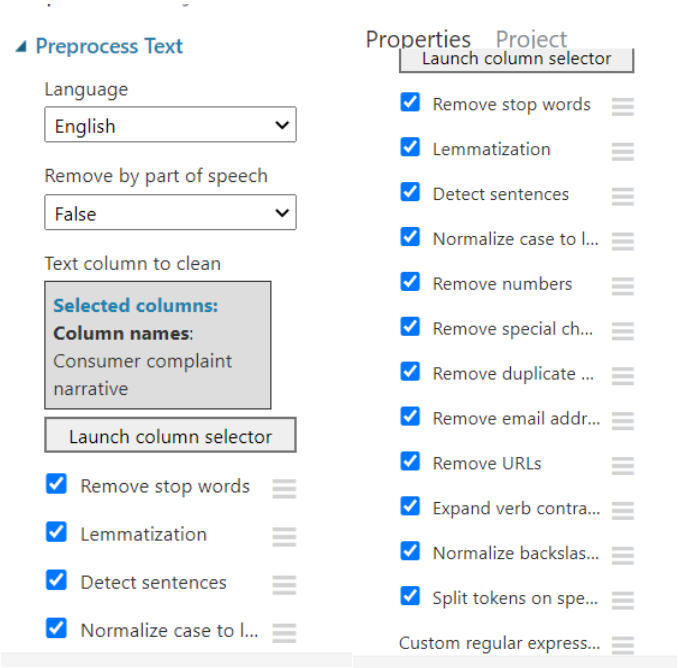


Text Analytics using Azure ML: -

Q6) Preprocess text data. Remove punctuations, stop words, lemmatization, etc. (2 points)

Ans.

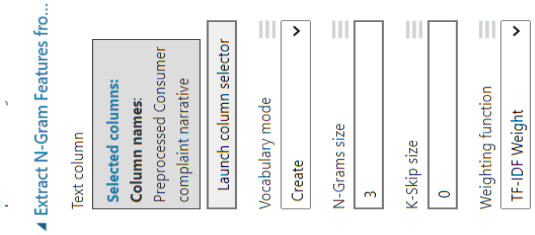




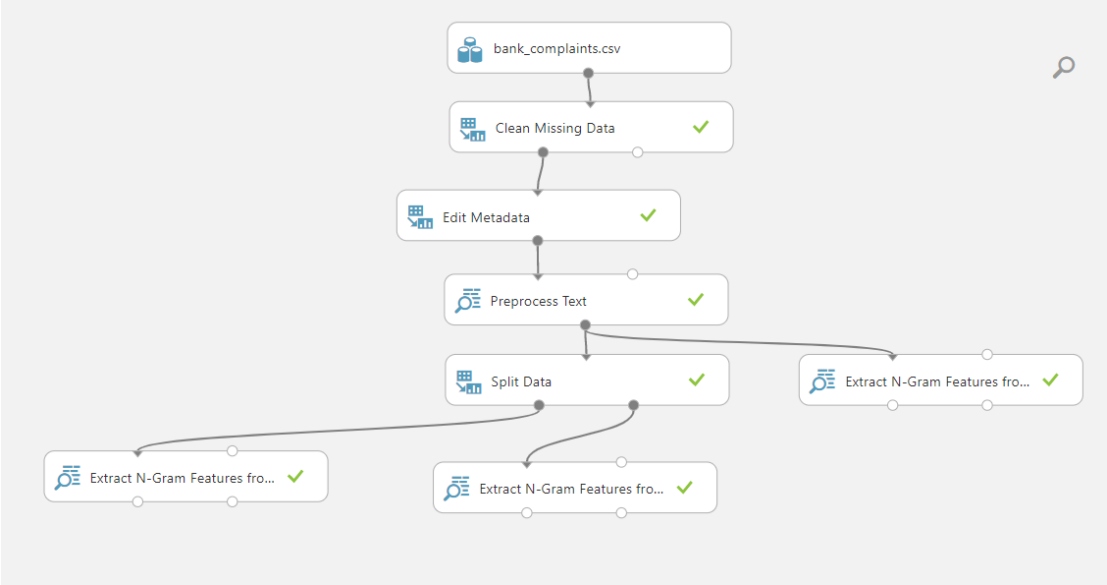
Preprocessed the data – comments/reviews from the customers

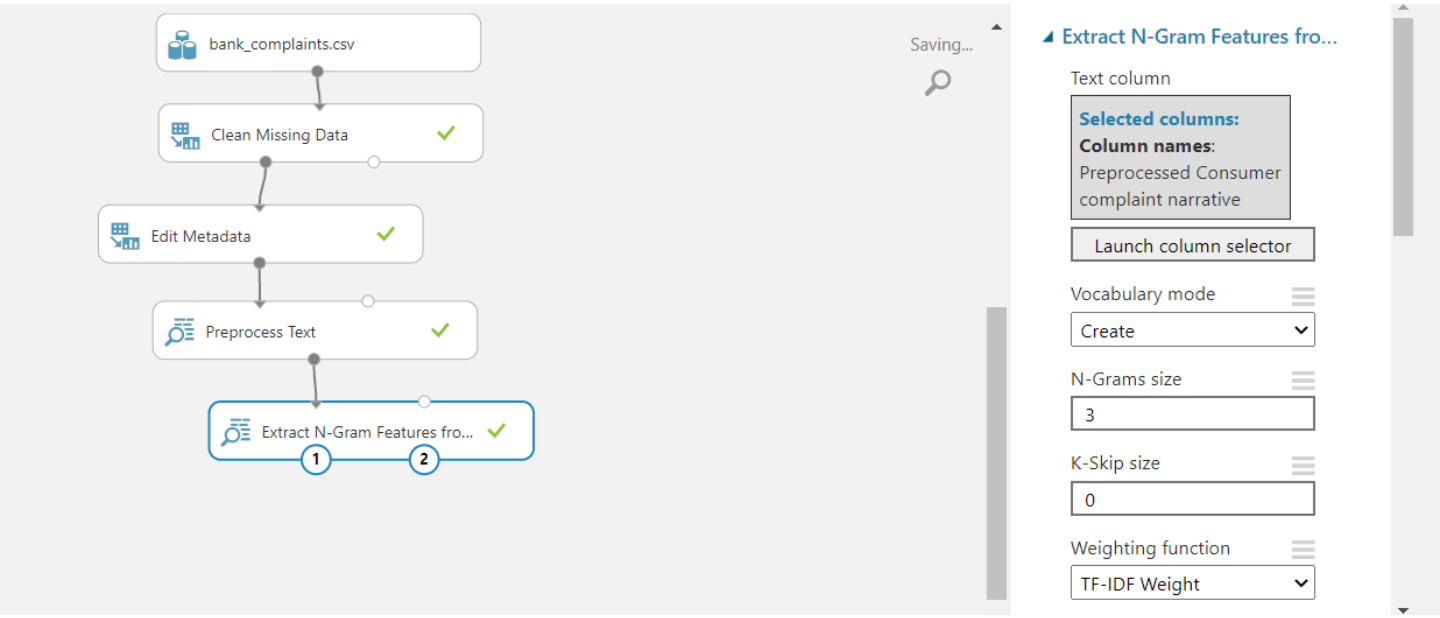
Q7) Extract N-gram features from the above preprocessed text and report the final number of columns. Report the size of the resulting vocabulary. (3 points)

Ans.

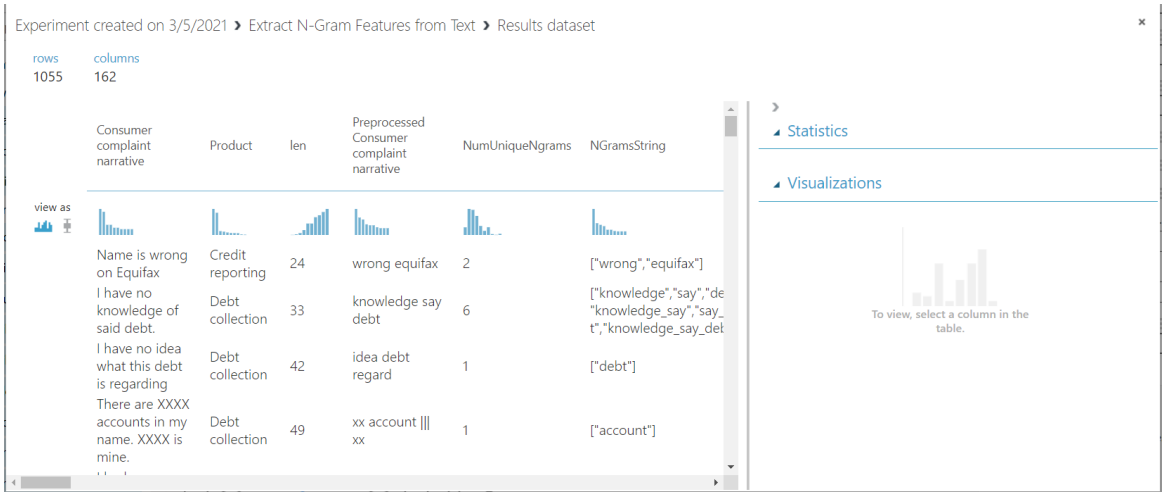


Using count based feature scring method.

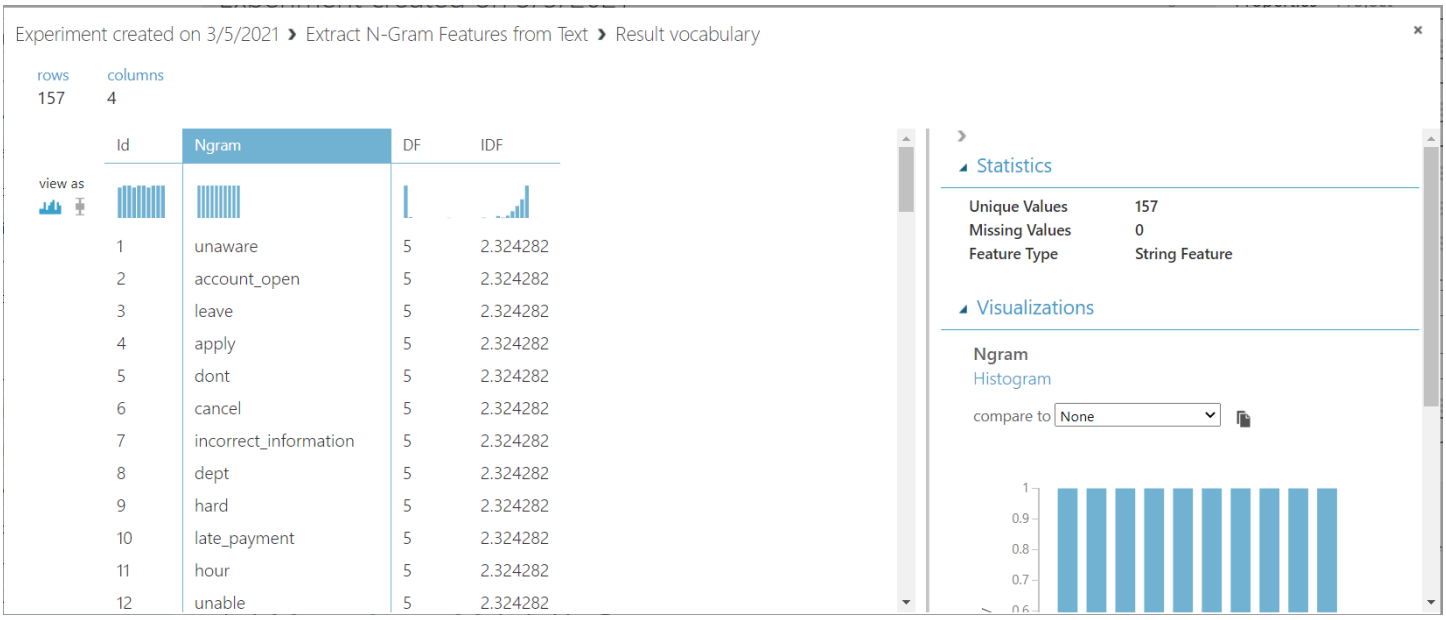




Final number of columns in the resulting dataset is 162.



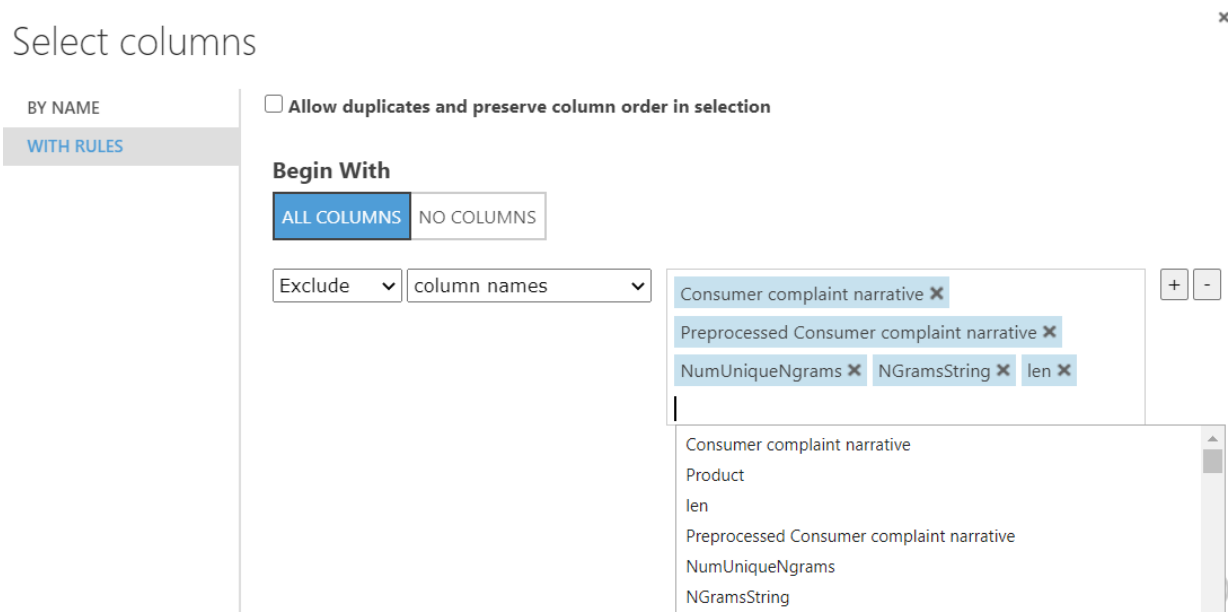
Resulting vocabulary is of 157 words.



Product class prediction using multi-class neural network: -

Q8) Select variables of interest for the model building process. (1 point)

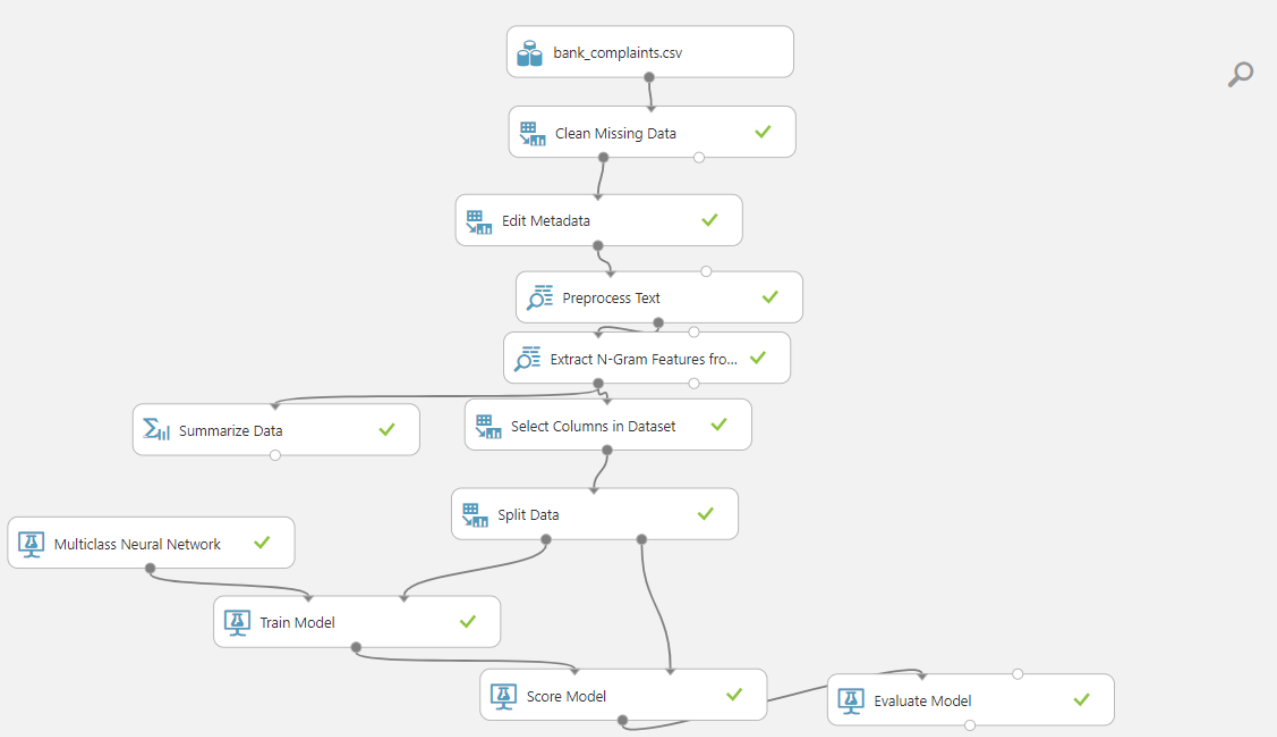
Ans.

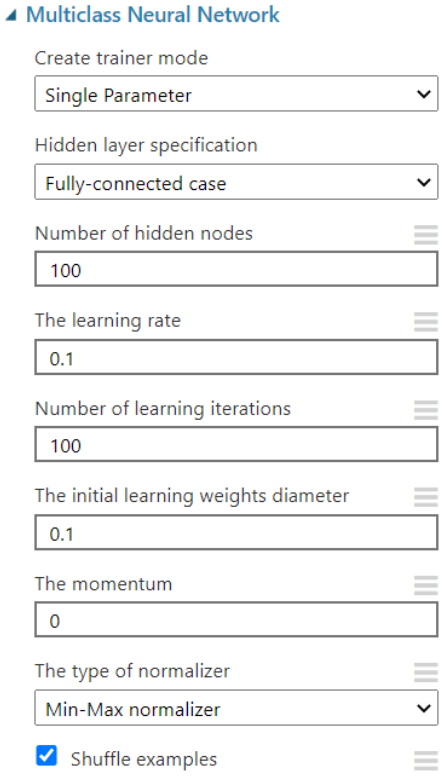


These are the features I think should be excluded for model training and testing.

Q9) Build a neural network model to predict the product name. (2 points)

Ans. Trained a multiclass neural network to predict the product name with 100 hidden layers and learning rate 0.1.



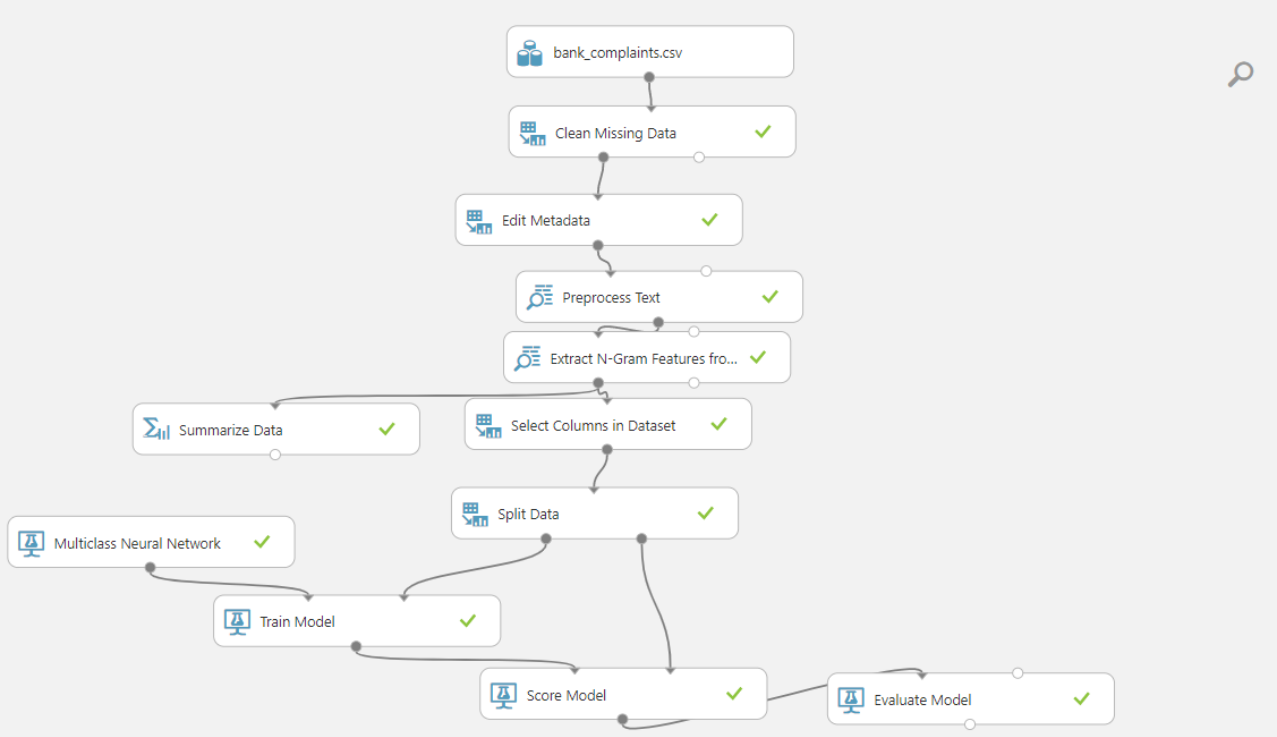


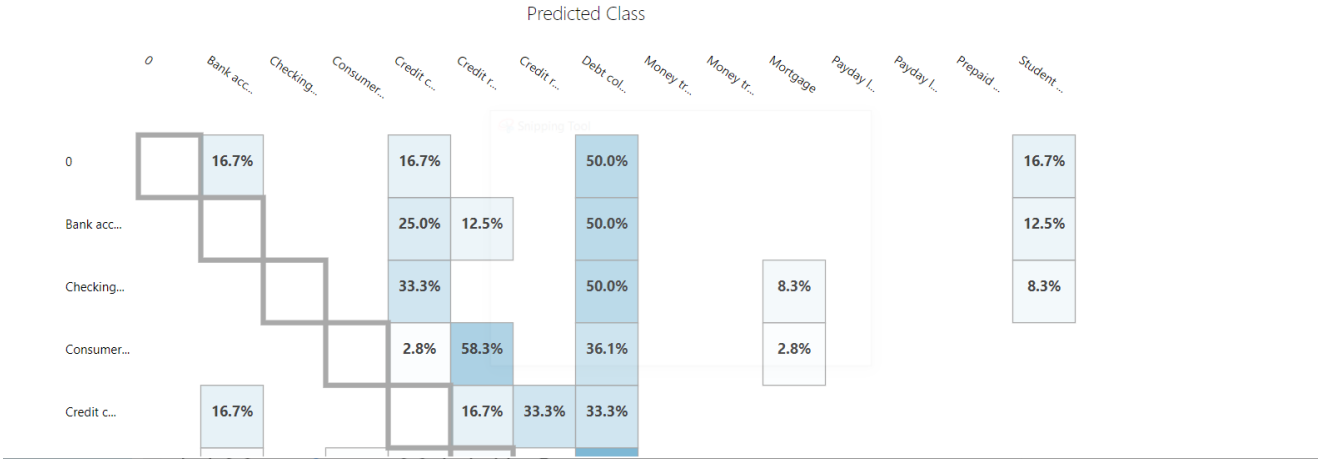
Q10) Evaluate the model and report the accuracy. (1 point)

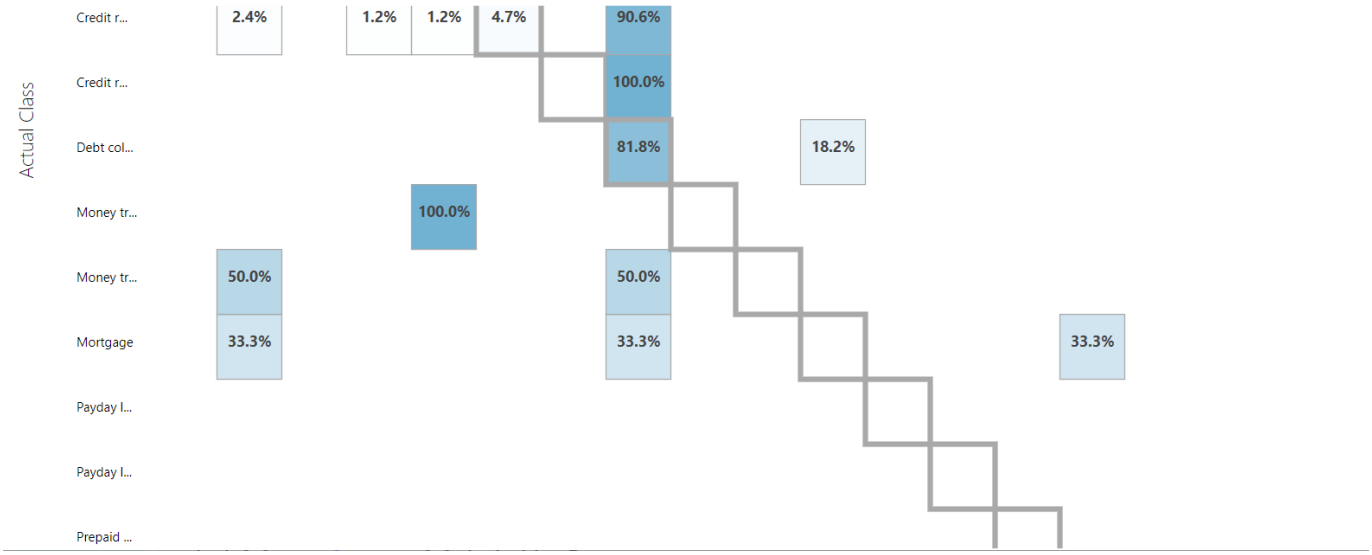
Ans. overall accuracy is 0.061905 with an average accuracy of 0.874921.

Metrics

Overall accuracy	0.061905
Average accuracy	0.874921
Micro-averaged precision	0.061905
Macro-averaged precision	NaN
Micro-averaged recall	0.061905
Macro-averaged recall	NaN







Q11) Write your observations and findings. (1 point)

Ans.

My observations are there are certain words which definitely point to the product that is complaint against like something related to credit return definitely indicates its related to debt collection. All these observations and findings are based on confusing matrix formed. Similarly money transferred to credit distribution so on so forth.