

Organizing Your Grievances

Classifying Multi-Class Consumer
Complaints using **Natural Language
Processing** (TFIDF/Word2Vec)



WHAT IS THIS ABOUT?

I am dissatisfied with the current outcome of a dispute that was initiated with Discover Card regarding a single transaction that occurred on XXXX/XXXX/2015 in the amount of {\$280.00}. I have corresponded with Discover Card at least four times since XXXX/XXXX/2015 (which I have enclosed as an attachment to this complaint). I believe that the credit card issuer has violated consumer protection laws by failing to implement the Special Rule for Credit Card Purchase protection despite overwhelming paperwork evidence submitted by me that shows the merchant has conducted business in bad faith less favorable to the consumer. I have sustained a monetary loss as a result of merchants bad faith and intent. I have patiently utilized the internal Discover Card dispute process over the past three months with the credit card issuer always favoring the merchant ; I have repeatedly submitted irrefutable paperwork evidence that has shown that the merchant has conducted business in bad faith. I have tried in good faith to address my complaint with the merchant and Discover Card but believe that I will not receive a favorable outcome.

READING EXERCISE

I am dissatisfied with the current outcome of a dispute that was initiated with Discover Card regarding a single transaction that occurred on XXXX/XXXX/2015 in the amount of {\$280.00}. I have corresponded with Discover Card at least four times since XXXX/XXXX/2015 (which I have enclosed as an attachment to this complaint). I believe that the credit card issuer has violated consumer protection laws by failing to implement the Special Rule for Credit Card Purchase protection despite overwhelming paperwork evidence submitted by me that shows the merchant has conducted business in bad faith less favorable to the consumer. I have sustained a monetary loss as a result of merchants bad faith and intent. I have patiently utilized the internal Discover Card dispute process over the past three months with the credit card issuer always favoring the merchant ; I have repeatedly submitted irrefutable paperwork evidence that has shown that the merchant has conducted business in bad faith. I have tried in good faith to address my complaint with the merchant and Discover Card but believe that I will not receive a favorable outcome.

**CREDIT
CARD**

WHAT ABOUT MORE?

I suffered a work-related injury in XXXX and am now permanently

XXXX. My income has dropped substantially in the past 5 years

am now

to XXXX

purchase

in the

Bank

made at le

work-related injury in XXXX and am now permanently

XXXX. My income has dropped substantially in the past 5 years

am now on a low fixed income. On XXXX separate occasions from XXXX

to XXXX I

purchase f

in the pro

Bank) card

made at le

work-related injury in XXXX and am now permanently

XXXX. My income has dropped substantially in the past 5 years and I

low fixed income. On XXXX separate occasions from XXXX

and zero income and had to use my Sams Club card to

d, thus running up debt out of necessity. I am currently

ss of attempting to pay down my Sams Club (Synchrony

and stopped using the card some time ago. I have always

t minimum payments and have never missed a payment.

my interest rate has been unilaterally raised three

past two years, and is now at 23.15 %. I called a Sams

rep today to file a complaint over the phone, because I

ng to be able to pay down this card when I am paying

00) a month in interest and can only afford to pay the

ent + {\$4.00} or {\$5.00} dollars. They would not work

ch I expected. In my opinion, Synchrony Bank is taking

I suffered a work-related injury in XXXX and am now permanently

XXXX. My income has dropped substantially in the past 5 years and I

am n

to X

purc

on a low fixed income. On XXXX separate occasions from XXXX

X I had zero income and had to use my Sams Club card to

I suffered a work-related injury in XXXX and am now permanently

XXXX. My income has dropped substantially in the past 5 years and I

income. On XXXX separate occasions from XXXX

income and had to use my Sams Club card to

running up debt out of necessity. I am currently

I suffered a work-related injury in XXXX and am now permanently

XXXX. My income has dropped substantially in the past 5 years and I

am now on a low fixed income. On XXXX separate occasions from XXXX

to XXXX I had zero income and had to use my Sams Club card to

purchase food, thus running up debt out of necessity

in the process of attempting to pay down my Sams Club

Bank) card, and stopped using the card some time ago

made at least minimum payments and have never missed

Despite this, my interest rate has been unilaterally raised three

times in the past two years, and is now at 23.15 %. I called a Sams

Club account rep today to file a complaint over the phone, because I

am never going to be able to pay down this card when I am paying

almost {\$50.00} a month in interest and can only afford to pay the

minimum payment + {\$4.00} or {\$5.00} dollars. They would not work

with me, which I expected. In my opinion, Synchrony Bank is taking

unfair advantage of recent interest rate hikes to go

especially those who are financially unable to make

PROBLEM STATEMENT

- Pressing and common problem in Customer Service Department in banks
- Thousands of complaints like these
- Nasty, unreadable wall-of-texts
- Complaints can be filed into wrong sections
- Difficulty in going through every single complaint and direct them to relevant departments to be dealt with



SO HOW DO WE DEAL WITH THIS ISSUE THEN?

PROJECT OBJECTIVE

APP that can **correctly classify** a customer
complaint according to product type

HOW THE APP WORKS

COMPLAINTS



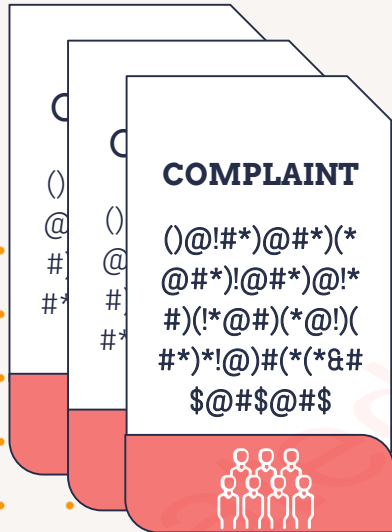
NLP

CLASSIFIER
MODEL

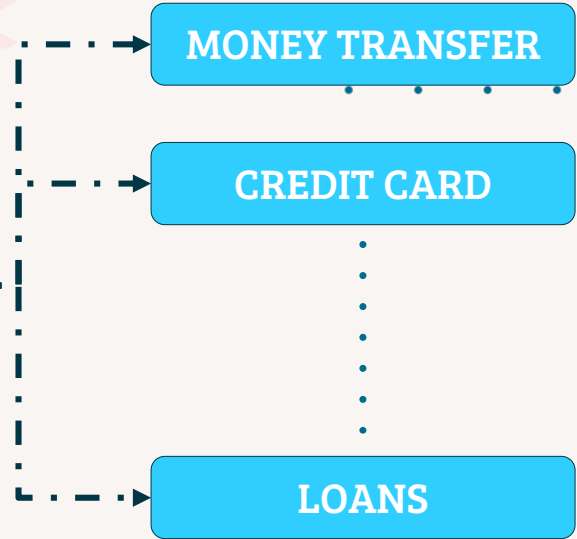


HOW THE APP WORKS

COMPLAINTS



NLP



APP BENEFITS

WORKLOAD REDUCTION



Free up man-power to be allocated on more important tasks

FASTER RESPONSE



App processes complaint instantly

ROUTINE TASK AUTOMATION



More efficiency in workflow

HOW IS THE MODEL BUILT?

1. DATA USED



Consumer Financial Protection Bureau

COMPLAINTS RECEIVED ABOUT FINANCIAL
PRODUCTS AND SERVICE





Consumer Financial Protection Bureau

COMPLAINTS RECEIVED ABOUT FINANCIAL
PRODUCTS AND SERVICE



30,000 COMPLAINTS



Consumer Financial Protection Bureau



COMPLAINTS RECEIVED ABOUT FINANCIAL
PRODUCTS AND SERVICE

30,000 COMPLAINTS

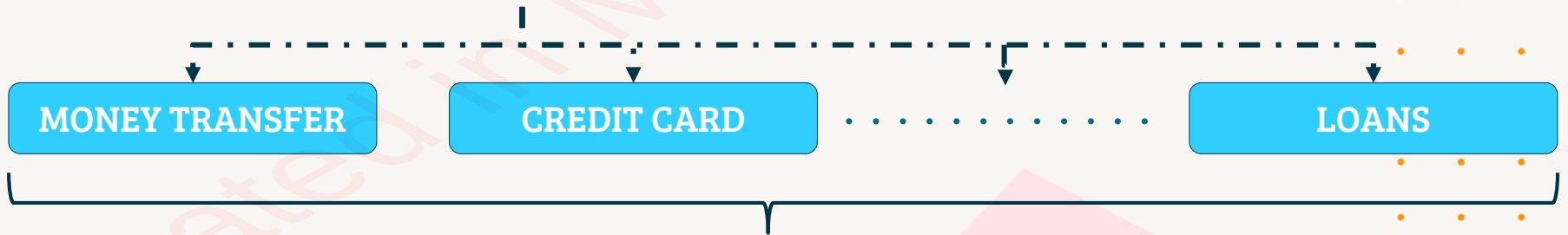


Consumer Financial Protection Bureau

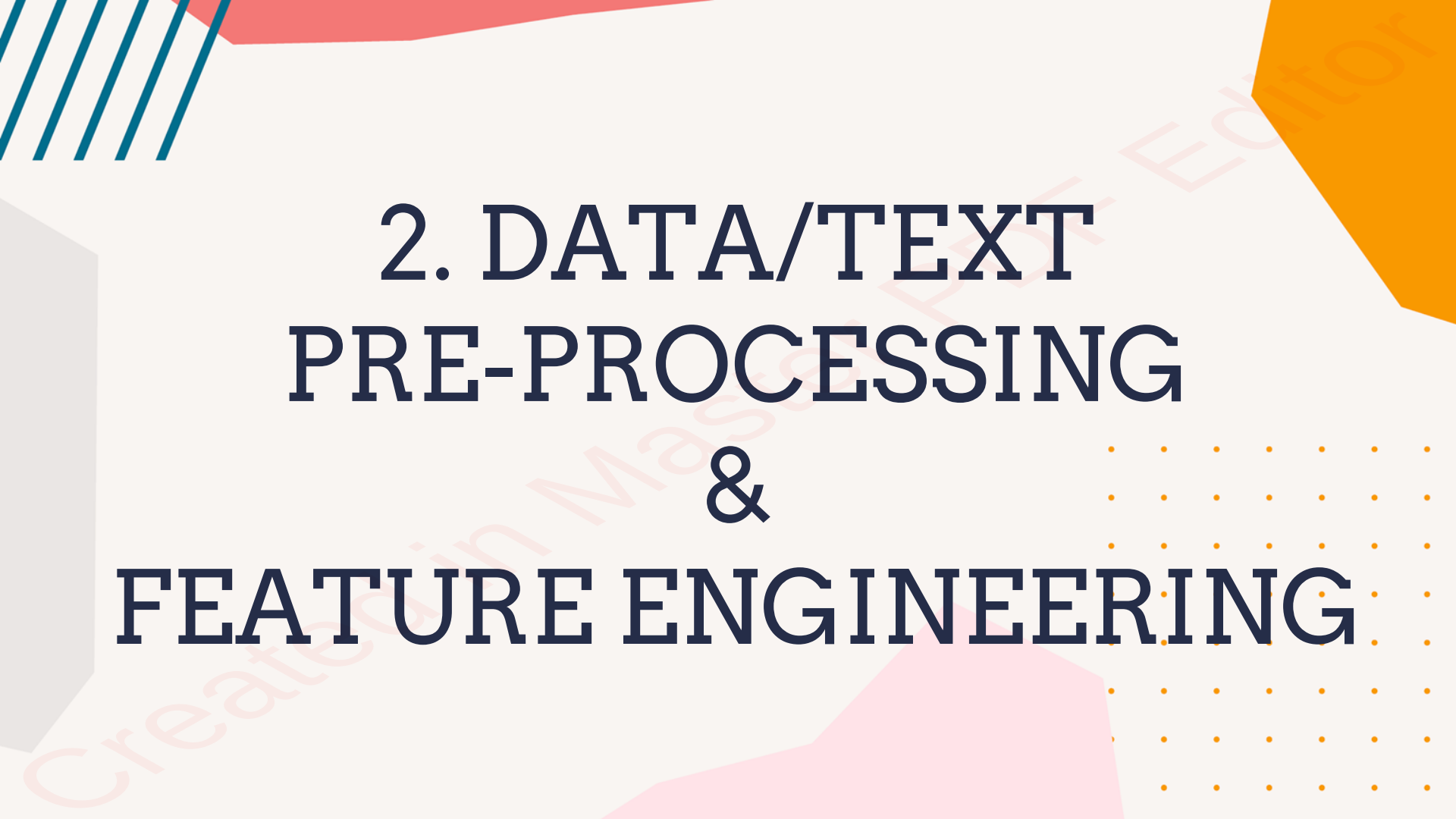


COMPLAINTS RECEIVED ABOUT FINANCIAL
PRODUCTS AND SERVICE

30,000 COMPLAINTS



10 CLASSES/PRODUCTS

The background features several geometric elements: a red polygon in the top left, a grey polygon on the left side, an orange polygon in the top right, and a pink polygon in the bottom center. There are also blue diagonal lines in the top left and a grid of orange dots in the bottom right.

2. DATA/TEXT PRE-PROCESSING & FEATURE ENGINEERING

FEATURE ENGINEERING

- Punctuation removal

I think Pair Programming is SO fun. I come to class feeling energized everyday and every unsolvable pair just pushes me to do better, broaden my horizons and helps keep my sanity intact.

FEATURE ENGINEERING

- Punctuation removal

I think Pair Programming is SO fun. I come to class feeling energized everyday and every unsolvable pair just pushes me to do better, broaden my horizons and helps keep my sanity intact.

FEATURE ENGINEERING

- Punctuation removal
- Stop-word removal

I think Pair Programming is SO fun. I come to class feeling energized everyday and every unsolvable pair just pushes me to do better, broaden my horizons and helps keep my sanity intact.

FEATURE ENGINEERING

- Punctuation removal
- Stop-word removal
- Lowercase

I think Pair Programming is SO fun. I come to class feeling energized everyday and every unsolvable pair just pushes me to do better, broaden my horizons and helps keep my sanity intact.

FEATURE ENGINEERING

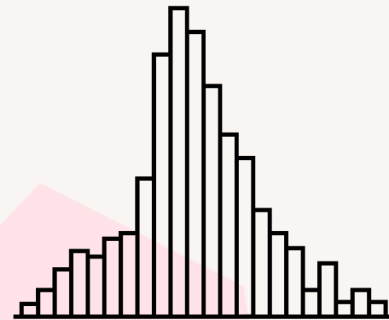
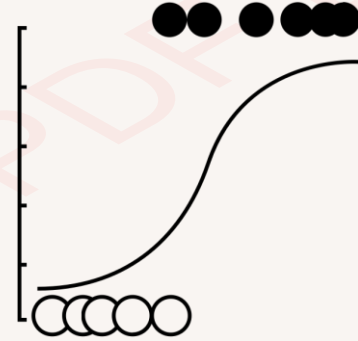
- Punctuation removal
- Tokenize
- Stop-word removal
- Ready for split for train/test/vectorization
- Lowercase

```
["pair", "programming", "fun", "class", "feeling", "energized",  
"everyday", "unsolvable", "pair", "pushes", "broaden", "horizons",  
"helps", "sanity", "intact"]
```


3. TRAINING METHODOLOGY

MODELS USED

- Multinomial Naïve Bayes
- Gaussian Naïve Bayes
- Logistic Regression
- Random Forest
- LinearSVC



CLASSIFIER EVALUATION METRIC

- Micro F1

CLASSIFIER EVALUATION METRIC

- Micro F1
- Measure of how well classifier gets it's predictions correct over entire dataset

CLASSIFIER EVALUATION METRIC

- Micro F1
- Measure of how well classifier gets its *predictions correct* over *entire dataset*

		Predicted			
		A	B	C	
True labels	A	2	2	0	4
	B	1	2	0	3
	C	0	0	3	3
		3	4	3	Total

CLASSIFIER EVALUATION METRIC

- Micro F1
- Measure of how well classifier gets its *predictions correct* over *entire dataset*
- Measure of classifier *overall accuracy*

		Predicted			
		A	B	C	
True labels	A	2	2	0	4
	B	1	2	0	3
	C	0	0	3	3
		3	4	3	Total

DATA SPLIT

60% Train Data

20% Holdout Data

20% Validation

Stratified 5 fold Cross-Validation

60% Train

20% Val

20% Test

TFIDF VECTORIZATION

Vectorize Text into Term-Document Matrix (1-gram)

Higher weightage to rare words

Words like "citi" for citi bank will hold more significance

60% Train Vec

20% Val Vec

20% Test Vec

TFIDF VECTORIZATION

Vectorize Text into Term-Document Matrix (1-gram)

Higher weightage to rare words

Words like "citi" for citi bank will hold more significance

60% Train Vec

20% Val Vec

20% Test Vec

4. CLASSIFIER SELECTION

TFIDF CROSS-VAL RESULTS

MODEL	MICRO F1
MULTINOMIALNB_TFIDF	0.69
GAUSSIANNB_TIDF	0.40
LOGISITC REG_TFIDF	0.83
RANDOMFOREST_TFIDF	0.44
LINEARSVC_TFIDF	0.82

60% Train Vec

20% Val Vec

20% Test Vec

TFIDF CROSS-VAL RESULTS

MODEL	MICRO F1
MULTINOMIALNB_TFIDF	0.69
GAUSSIANNB_TIDF	0.40
LOGISITC REG_TFIDF	0.83
RANDOMFOREST_TFIDF	0.44
LINEARSVC_TFIDF	0.82

83%

60% Train Vec

20% Val Vec

20% Test Vec

TFIDF CROSS-VAL RESULTS

MODEL	MICRO F1
MULTINOMIALNB_TFIDF	0.69
GAUSSIANNB_TIDF	0.40
LOGISITC REG_TFIDF	0.83
RANDOMFOREST_TFIDF	0.44
LINEARSVC_TFIDF	0.82

83%

But can we do better?

60% Train Vec

20% Val Vec

20% Test Vec

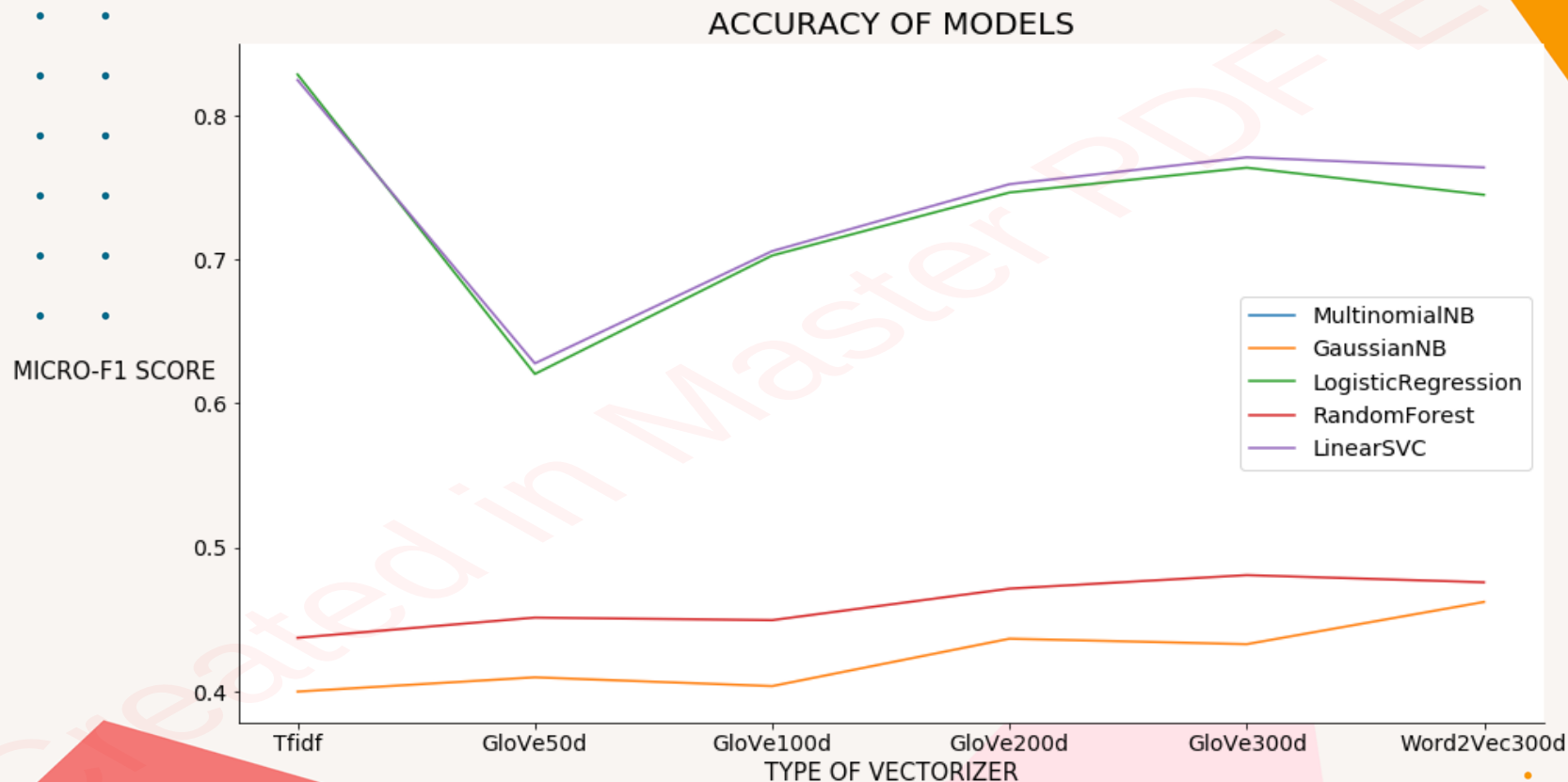
GloVe & Word2Vec



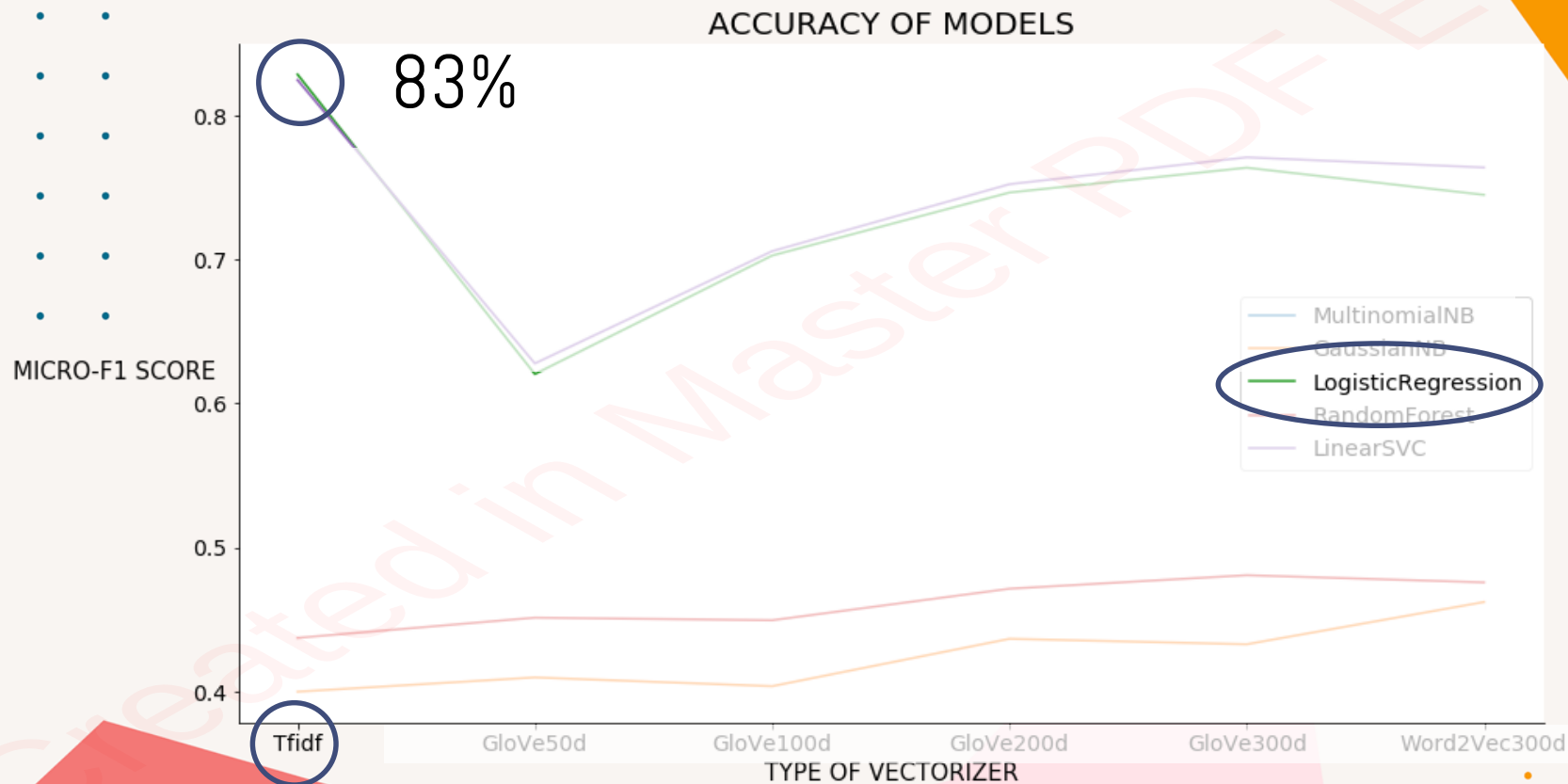
Stanford



CROSS-VALIDATION RESULTS



CROSS-VALIDATION RESULTS



FINAL TRAIN/TEST

LOGISTIC REGRESSION CLASSIFIER PERFORMANCE ON UNSEEN DATA

80% Train Vec

20% Test Vec

RESULTS

83%



Model correctly classifies
83% of the time

The background features a light cream color with several abstract geometric elements: a red shape at the top left, a grey shape on the left, a yellow shape at the top right, and a pink shape at the bottom. A grid of small orange dots is located in the bottom right corner. A faint, diagonal watermark reading 'Created in PDF Editor' is visible across the center.

FLASK DEMONSTRATION

SAMPLE COMPLAINT

I am dissatisfied with the current outcome of a dispute that was initiated with Discover Card regarding a single transaction that occurred on XXXX/XXXX/2015 in the amount of {\$280.00}. I have corresponded with Discover Card at least four times since XXXX/XXXX/2015 (which I have enclosed as an attachment to this complaint). I believe that the credit card issuer has violated consumer protection laws by failing to implement the Special Rule for Credit Card Purchase protection despite overwhelming paperwork evidence submitted by me that shows the merchant has conducted business in bad faith less favorable to the consumer. I have sustained a monetary loss as a result of merchants bad faith and intent. I have patiently utilized the internal Discover Card dispute process over the past three months with the credit card issuer always favoring the merchant ; I have repeatedly submitted irrefutable paperwork evidence that has shown that the merchant has conducted business in bad faith. I have tried in good faith to address my complaint with the merchant and Discover Card but believe that I will not receive a favorable outcome.

**CREDIT
CARD**

Consumer Complaints Classifier

Good-day Sir/Ma'm. This is your friendly neighborhood Complaint Classifier!

How may I address your complaint today?

Submit

I think you are concerned about...

[illegible]

HOW ABOUT A MORE DIFFICULT COMPLAINT?

Omg where is my money?????? AOSINIONSAD WHY DID YOU GUYS TAKE MY MONEY AWAY. SAOIDNSIOADNOIAODNNIOASDNSADNOSDNOASDN

I TRANSFERRED IT LAST NIGHT AND I WOKE UP TO NOTHING IN!!!!!! MY BANK ACCOUNT. HELP PLEASE!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

I NEED THE MONEY OR ELSE I WILL BE HUNTED BY LOAN SHARKS!!!!!!!!!!

YOU *(&\$) PEOPLE HAVE NO RIGHT TO DO THIS TO ME!!!! I NEED MY MONEY!!!!

HOW ABOUT A MORE DIFFICULT COMPLAINT?

Omg where is my money????? AOSINIONSAD WHY DID YOU GUYS TAKE MY
MONEY AWAY. SAOIDNSIOADNOIAODNNIOASDNSADNOSDNOASDN

I TRANSFERRED IT LAST NIGHT AND I WOKE UP TO NOTHING IN!!!!!! MY
BANK ACCOUNT. HELP PLEASE!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

I NEED THE MONEY OR ELSE I WILL BE HUNTED BY LOAN SHARKS!!!!!!!!!!

YOU *(#&\$) PEOPLE HAVE NO RIGHT TO DO THIS TO ME!!!! I NEED MY
MONEY!!!!

Messy

HOW ABOUT A MORE DIFFICULT COMPLAINT?

Omg where is my money????? AOSINIONSAD WHY DID YOU GUYS TAKE MY MONEY AWAY. SAOIDNSIOADNOIAODNNIOASDNSADNOSDNOASDN

I TRANSFERRED IT LAST NIGHT AND I WOKE UP TO NOTHING IN!!!!!! MY BANK ACCOUNT. HELP PLEASE!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

I NEED THE MONEY OR ELSE I WILL BE HUNTED BY LOAN SHARKS!!!!!!!!!!

YOU * (#&\$) PEOPLE HAVE NO RIGHT TO DO THIS TO ME!!!! I NEED MY MONEY!!!!

Messy

**What is this
complaint
about?**

Consumer Complaints Classifier

Good-day Sir/Ma'm. This is your friendly neighborhood Complaint Classifier!

How may I address your complaint today?

Submit

I think you are concerned about...

[illegible]

CORRECT PREDICTION

Omg where is my money????? AOSINIONSAD WHY DID YOU GUYS TAKE MY MONEY AWAY. SAOIDNSIOADNOIAODNNIOASDNSADNOSDNOASDN

I TRANSFERRED IT LAST NIGHT AND I WOKE UP TO NOTHING IN!!!!!! MY BANK ACCOUNT. HELP PLEASE!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

I NEED THE MONEY OR ELSE I WILL BE HUNTED BY LOAN SHARKS!!!!!!!!!!

YOU *(&\$) PEOPLE HAVE NO RIGHT TO DO THIS TO ME!!!! I NEED MY MONEY!!!!

Money Transfer



PRATICAL USES OF APP

INSTANTLY CLASSIFY A COMPLAINT

Faster response time to customers

FREE UP MANPOWER ON ROUTINE TASKS

Greater efficiency

ORGANIZE COMPLAINTS WITHOUT ACTUALLY READING THEM

Reduce strain on customer service department



FUTURE WORK



TOPIC SUMMARIZATION FUNCTIONALITY


Provide short abstract of each complaint

COMPLAINT PRIORITIZATION

Sentiment analysis

Prioritize angrier-sounding complaints

Detect vulgarities



TFIDF VECTORIZATION

- Tokenize Text into Term-Document Matrix using 1-gram

*TFIDF score for term i in document $j = TF(i, j) * IDF(i)$*

where

IDF = Inverse Document Frequency

TF = Term Frequency

$$TF(i, j) = \frac{\text{Term } i \text{ frequency in document } j}{\text{Total words in document } j}$$

$$IDF(i) = \log_2 \left(\frac{\text{Total documents}}{\text{documents with term } i} \right)$$

and

t = Term

j = Document

