Customer Service Presentation

By Ritvik Iyer, James Jung, Anurag Aiyer

GitHub Link: https://github.com/ritvik-iyer/microsoftxpieDatathon

Customer Service Data Features

SupportTicket: unique ID for support ticket

CustomerID: unique ID when customer submits ticket

DataCreated: creation of ticket

DataCompleted: end of ticket

Escalated: whether urgency of ticket is immediate

Theme: category of support ticket

Text: customer complaint

Sentiment: Emotion of text

Key Phrases: most important bodies of text

Initial Questions

a. How long does it take for a support ticket to be resolved?

a. Potential fraudulent/spam support tickets

a. What factors influence a ticket to be escalated?

a. Classifier which predicts theme based on text

Initial Observations

- **MOST** number of tickets come from the following themes: **speed, price, & features**

- On average, **escalated** tickets take **10 days longer** to process than **non-escalated** tickets

- **Reliability**, **security** and **other-themed tickets** take the longest to

Process Time

Escalated

0	30.993033
1	41.277083

Process Time

Theme	
design	4.132479
features	29.637249
other	38.549451
price	35.331450
reliability	49.712366
security	44.244344
services	27.396008
speed	30.245066
support	24.392857

Settling on one question

Given the customer text input and theme of ticket, predict its estimated processing time.

Our Discoveries

- About **83%** of support tickets are **neutral** and **negative**
- Average Processing Time when tickets are **NOT** escalated: **Neutral** tickets takes the most time
- Average Processing Time when tickets are escalated: **Negative** tickets take the most time.

Most Common Words for Speed Theme

Most Common Words for Price Theme

Most Common Words for Features Theme

Feature Engineering Part 1

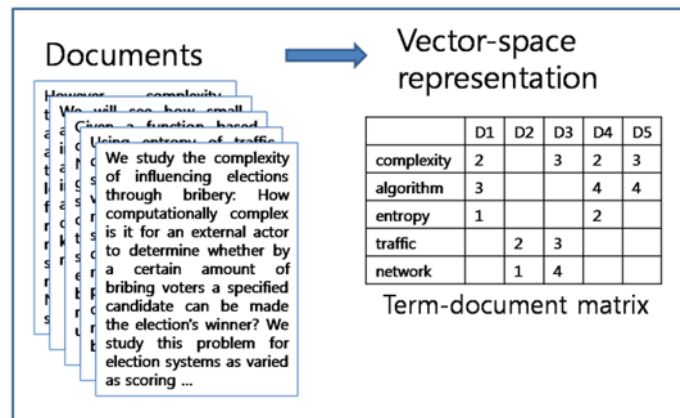1) Needed to clean the data heavily:
   a) The Process Times (Date Completed - Date Created) were sometimes **negative**
   b) Replaced negative values with average of positive times grouped by theme and escalated

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$tf_{i,j}$ = number of occurrences of i in j
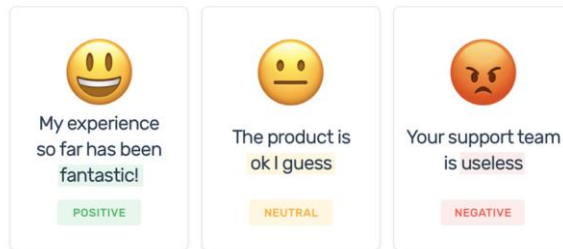
df_i = number of documents containing i

N = total number of documents

Documents → Vector-space representation

We study the complexity of influencing elections through bribery: How computationally complex is it for an external actor to determine whether by a certain amount of bribing voters a specified candidate can be made the election's winner? We study this problem for election systems as varied as scoring ...

	D1	D2	D3	D4	D5
complexity	2		3	2	3
algorithm	3			4	4
entropy	1			2	
traffic		2	3		
network		1	4		

Term-document matrix

Feature Engineering Part 2

1) Using Azure Text Analysis API to process text and extract **sentiment** and key **phrases**
   a) Used sentiment_analysis_example and extract_key_phrases functions to retrieve desired information
   b) Looked at part of speech & text ranking
2) Put all this information and cleaned up data further


Sentiment Analysis

😃
My experience
so far has been
fantastic!
POSITIVE

😐
The product is
ok I guess
NEUTRAL

😡
Your support team
is useless
NEGATIVE

Predicting Process Times Part 1

Linear

	Process Time	Predicted Process Time	Difference
204	4.000000	-7.647388e+14	7.647388e+14
266	4.000000	-2.937085e+14	2.937085e+14
82	26.750000	-2.813502e+14	2.813502e+14
254	245.000000	-1.397741e+14	1.397741e+14
231	30.000000	-1.163420e+14	1.163420e+14
...	...	...	...
146	37.755556	1.395899e+14	-1.395899e+14
25	1.000000	1.549951e+14	-1.549951e+14
5	6.000000	2.346103e+14	-2.346103e+14
114	6.000000	2.568234e+14	-2.568234e+14
275	0.000000	2.906664e+14	-2.906664e+14

Poisson

	Process Time	Predicted Process Time	Difference
254	245.0	20.907250	224.092750
180	273.0	65.125458	207.874542
148	273.0	65.454739	207.545261
3	273.0	68.817809	204.182191
209	212.0	32.506238	179.493762
...	...	...	...
314	4.0	51.295855	-47.295855
118	9.0	58.068124	-49.068124
12	1.0	51.175799	-50.175799
322	1.0	54.912387	-53.912387
15	2.0	58.878246	-56.878246

Predicting Process Times Part 2

Distribution of Differences in Predicted Value for Linear Regression
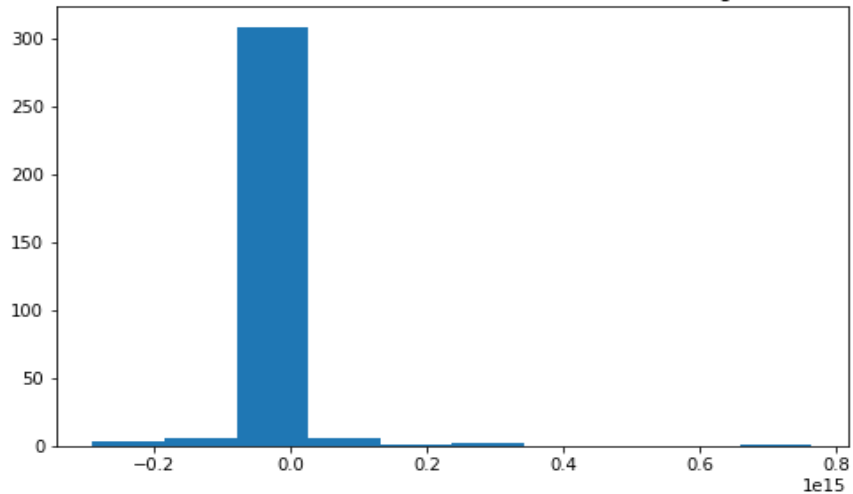
Distribution of Differences in Predicted Value for Poisson Regression