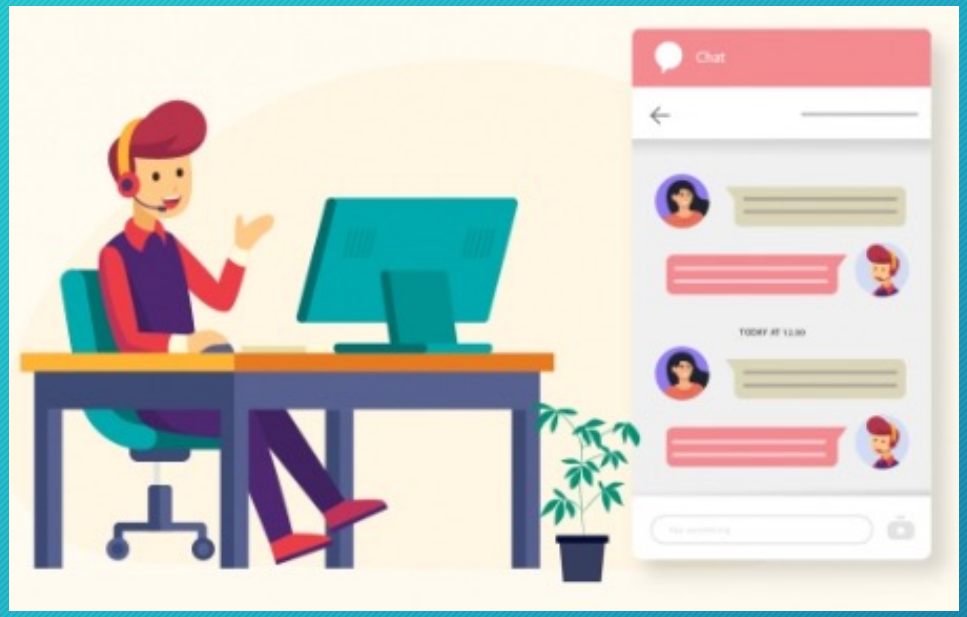
City of El Paso’s Ticketing System Classification using NLP methods

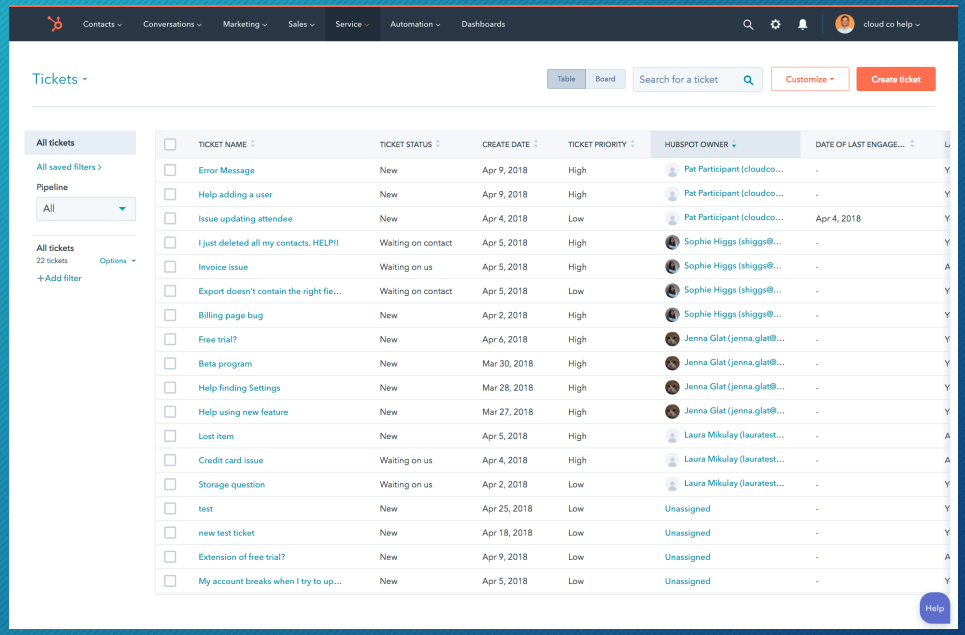
Outline

- Background
- Problem
- Exploratory data analysis
- Neural network model
- Results & conclusion
- Future work

Background

- City Departments such as Fire Department or Libraries can use ticketing system to request help to IT department. Examples include trouble shooting or equipment replacement
- IT expert reads request and manually assigns to correct IT team
- Want to create a model that classifies tickets to save IT expert some time





Problem

- Can we classify incoming tickets and assign them to the correct support IT team using Natural Language Processing?

Datasets origin

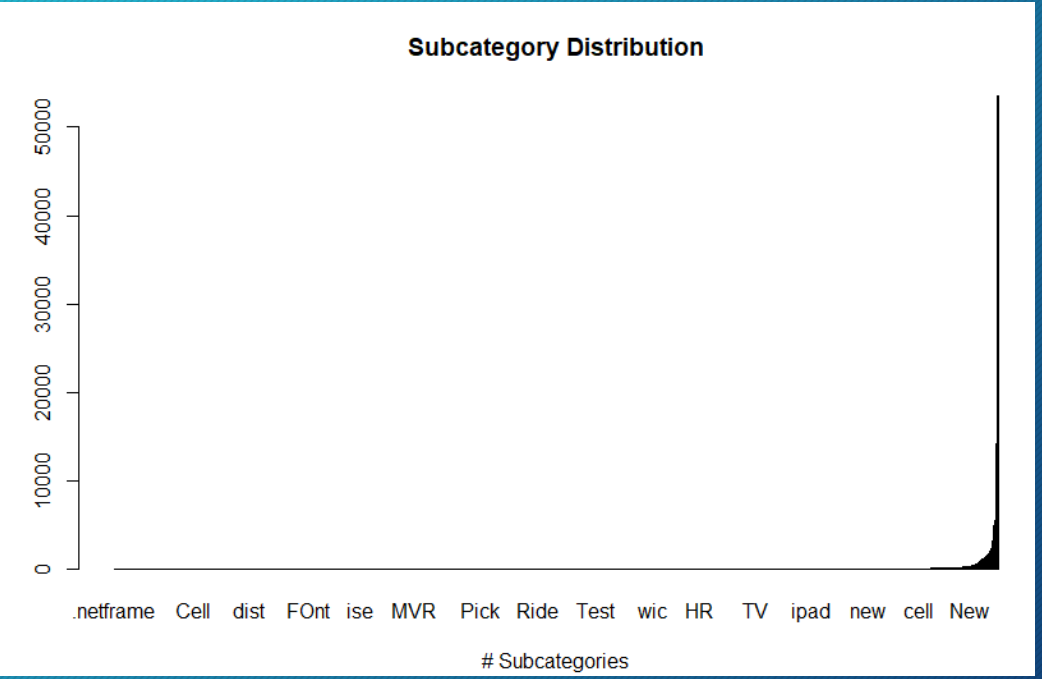
- Cherwell ticketing system dataset:

  Provided by City IT, no private fields to preserve Privacy.

Incident II	Priority	Description	CustomerDepartment	Owned By Team	Service	Category	Subcategory
472861	3	PLEASE NOTE THAT WE'RE ADDING NUMBERS ON RECOVERIES. CAN WE ADD A COLUM	City Manager	Web Administration	Web Services	Information Update	Content
472862	3	Work on Teams virtual backgrounds. Finish out Phase 1 timeline for SilverStripe city \	Information Technology	Web Administration	Web Services	Information Update	Content
472863	3	Assistance with reinstalling e-mail archives.\r\n \r\nMiranda, Luis	Public Health	Help Desk	Remote Support Services	Software Support	Troubleshooting
472864	3	OMB review meeting of 3 pending encumbrance issues.\r\n \r\nTaking over the last	Information Technology	Applications	Enterprise Apps	PeopleSoft Financials	Technical Support
472865	5	Please assist with configuring the new computers at George Perry testing site for the	Information Technology	Network Infrastructure	Network Services	Network Connectivity	Modify Service
472866	2	Teams call: plan for new featureclasses from testing database, new feature services,	Information Technology	GIS	GIS Services	Web Viewer	Update
472867	3	Microsoft SQL 2008 server migration to Microsoft SQL 2012.\r\n \r\n*SSRS ACCELA cor	Information Technology	Applications	Enterprise Apps	Accela	Accela Technical Support
472868	2	Teams call: new web map for covid dashboard	Information Technology	GIS	GIS Services	Web Viewer	Update
472869	5	Meeting with the Zoo to assist with failed survey\r\nAssiting Angel with research	Information Technology	Applications	IT General	General	General
472870	3	Current Laptop: Dell XPS 9350\r\nService Tag:F7KBPF2\r\nCity Tag: 2395550\r\nDepar	Information Technology	Support Services	Onsite Support Services	Computer/Tablets	Setup/Relocation
472871	3	Issue: Phoenix Expungement - receives unhandled error. See attachment.\r\nCaller:	Police	Applications	Non-Enterprise Apps	Application Support	Technical Support
472873	4	User reporting not able to print from FS19-02	Fire	Support Services	Remote Support Services	Software Support	Papercut
472874	5	Please assist with printer at George perry testing site.	Information Technology	Network Infrastructure	Network Services	Network Connectivity	Modify Service
472875	2	[Performance Office - The 500]\r\nGood afternoon team, Could you please include th	Information Technology	Web Administration	Web Services	Electronic Form	Form Modification
472876	3	Research PS Script to find H drive folder size	Information Technology	Support Services	Remote Support Services	Software Support	Troubleshooting
472878	3	post meeting notices	Municipal Court	Web Administration	Web Services	Information Update	Content

Exploratory data analysis (EDA)

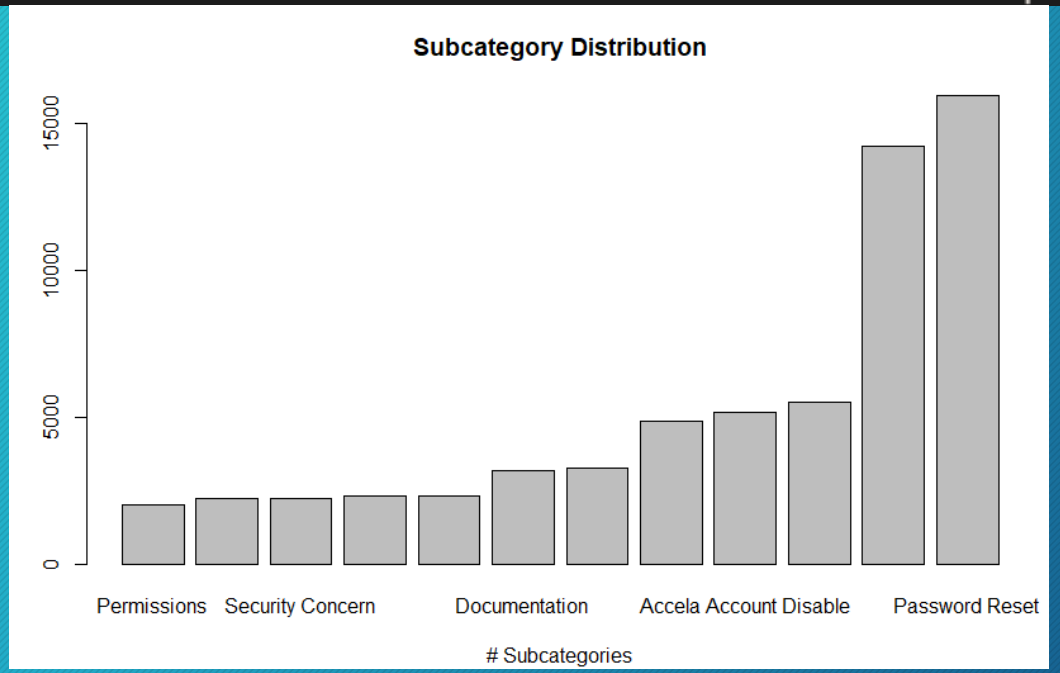
- Original dataset:
```
> glimpse(df)
Rows: 230,805
Columns: 23
```
- Predictor variable is Description and Target variable is Subcategory.
- Unique values for Subcategory: 2146.



Exploratory data analysis (EDA) – Cont.


```
  Accela Account Disable              Badge Access                   Content     Creation/Modification
                    5182                      4859                      5497                      2343
           Documentation       Employee Separation Installation/Modification           Password Reset
                    3204                     14223                      3271                     15952
             Permissions          Security Concern            Service Down           Unlock Account
                    2021                      2241                      2340                      2223
> unique(df2$Subcategory)
 [1] "Password Reset"           "Documentation"            "Permissions"              "Installation/Modification"
 [5] "Content"                  "Service Down"             "Security Concern"         "Badge Access"
 [9] "Creation/Modification"    "Unlock Account"           "Accela Account Disable"   "Employee Separation"
```

12 most common Subcategory Values:



Neural network model

Target variable, Subcategory, needed to be converted to numeric value and represented as a matrix-like data structure to use it as output nodes in the NN model.

```
> y_test
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
 [1,]    0    0    0    0    0    0    0    1    0     0     0     0
 [2,]    0    0    0    0    0    0    0    1    0     0     0     0
 [3,]    0    0    0    0    0    0    1    0    0     0     0     0
 [4,]    0    0    0    0    0    0    0    0    0     0     1     0
 [5,]    0    0    0    0    0    0    0    0    0     1     0     0
 [6,]    0    0    0    0    0    0    0    1    0     0     0     0
 [7,]    0    0    0    0    0    0    1    0    0     0     0     0
 [8,]    0    0    1    0    0    0    0    0    0     0     0     0
 [9,]    0    0    0    0    0    0    0    1    0     0     0     0
[10,]    0    0    0    0    0    0    1    0    0     0     0     0
[11,]    0    0    1    0    0    0    0    0    0     0     0     0
[12,]    0    0    0    0    0    0    0    1    0     0     0     0
[13,]    0    0    0    0    0    0    0    1    0     0     0     0
[14,]    0    0    1    0    0    0    0    0    0     0     0     0
[15,]    0    0    0    0    0    0    0    0    0     0     0     1
[16,]    0    0    0    0    0    0    0    1    0     0     0     0
```

Neural network model – Cont.

Predictor variable Description Remarks:

- Free-text value input by user.
- Removed stop words, e.g., “the”. “a”, “is”, “are”.
- Identified top 1000 most used words
  - “Please reset my password, I don’t remember it” and “My password has expired, I need to reset it”
  - “My computer mouse is broken, please replace it” and “A mouse made a mess in the kitchen, please replace utensils”

Neural network model – Cont.

- Assign an integer to each top 1000 words
- Convert Descriptions to integers
- Padded sequences with zeros to make Descriptions same length

```
[[1]]
[1]  26  63  14   3  72 152  44  38

[[2]]
 [1]  83 491  37 662 898  14  19 574 462 879   3  37   2  20  33  52 462 412  37  20  69 341

[[3]]
 [1] 922 108 239  83 491  37 662 898  14  19 574 462 879   3  37   2  20  33  52 462 412  37  20  69 341

[[4]]
 [1] 663 542  32 943 437 649  27 839 131 397 495 157 141  16   3  74 839 383   7 161 177  75 543  76 189  45 175  89 649
[30] 943 437 302 799 443 302 119

[[5]]
[1] 26 63 14  3 72 44 38

[[6]]
[1] 188 324
```

Neural network model – Cont.

Training the model:

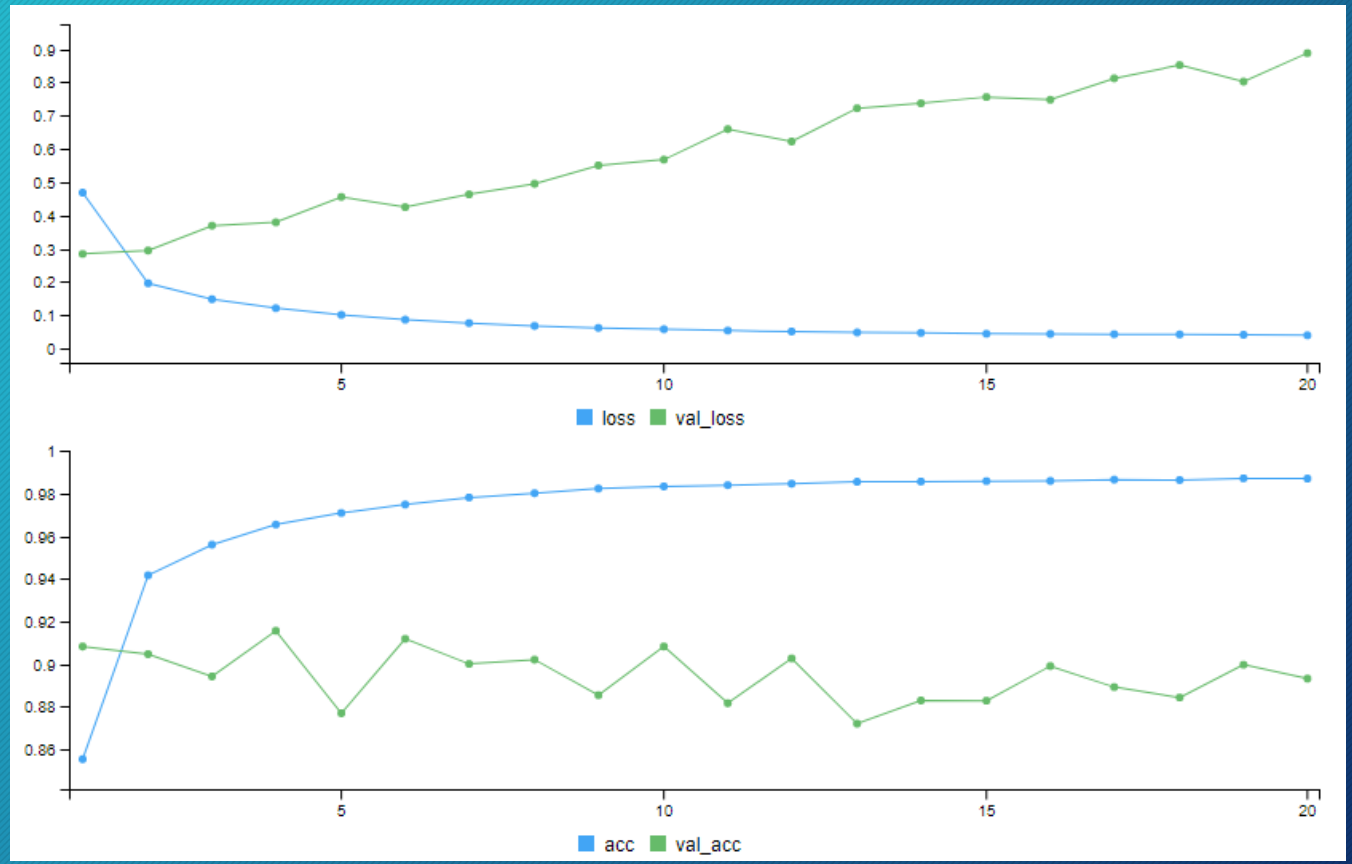
- Accuracy 97%

```
Overall Statistics

               Accuracy : 0.9707
                 95% CI : (0.9692, 0.9721)
    No Information Rate : 0.2522
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9657

 Mcnemar's Test P-Value : NA
```



Neural network model – Cont.

Testing the model:

- Accuracy 90%

```
Overall Statistics

               Accuracy : 0.9098
                 95% CI : (0.9047, 0.9147)
    No Information Rate : 0.2503
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.8942
```

Results and conclusion

- Neural network accuracy – **90%**
- Neural network classified our target data well.

Future work

- Try other approaches such as Support Vector Machines.
- Research and apply advanced natural language processing techniques.
- Apply text-embeddings process to compute useful statistics such as word proximity and synonym distinction.