

IT Case Data Analyzer

Requirements and Specification Document

Ray Smith, Josh Duchene, Francesca Tyler, Matthew Duff
Nathan Lenar, Daniel Owens, Nash Wellington

<2019-03-29>, version 3.0

Project Abstract

The IT Case Data Analyzer will drastically help the UW-Madison DoIT tech store in their post-processing of cases. Currently, Cherwell, the program that DoIT uses for categorizing cases, outputs data with broadly grouped cases into a hard to read .csv file. This data analyzer aims to take data outputted by Cherwell and classify these cases into more specific categories. This will minimize manual sorting that is currently done by tech store employees and will speed up the post-processing procedure.

Document Revision History

Rev. 1.0 <2019-02-16>: initial version

Rev. 2.0 <2019-03-29>: updates from iteration 1, changes to testing, changes to structure

Rev. 3.0 <2019-04-12>: updates from iteration 2

Customer

The client for this application is the Department of Information Services for UW-Madison aka DoIT. The main person within DoIT we expect to use this software is Patrick Johnson, the supervisor of the tech store, but anybody who works for DoIT would be able to use it. Currently, Mr. Johnson and other techstore employees organize many cases sent in from customers by hand. This is a painstaking task that could be made much easier by our program. Because we will be training the neural net specifically for DoIT, this software won't really be all that useful for people outside of DoIT unless some other IT department uses the exact same classification scheme. However, if another potential customer has .txt files that contain the the input keywords and the output categories, we could use those to retrain the neural network to work for them. It would take some work by developers to make sure it all worked smoothly, but it would be mostly just small tweaks. Such customers would include other departments that spend a lot of time classifying cases by hand.

Competitive Landscape

The function of our neural network is highly specialized and thus there's not a lot of competition for that half of the project. Currently our customer has to manually assign miscellaneous cases to their respective categories. An estimated 40-60 cases per hour can be organized manually by a competent staff member. If the staff member is paid \$10/hour as a low estimate, it could cost the customer anywhere from \$100 to \$150 per 600 cases that are not categorized based on labor fees.

The other side of our application outputs graphical representations of the data. Our "competitors" would be data visualization and organizational softwares such as Excel and Tableau. We don't consider these to be our competitors as we hope to have our software interface with these other applications.

User Requirements

- **Categorize New Cases (Top Priority)**
 - Import raw csv file
 - Neural Network will then be used to categorize the cases
 - Will flag cases where label's confidence level is low
 - Time permitting, would flag cases with multiple issues
 - Export categorized cases in new csv file
 - Export quick analysis of cases in pdf file
 - Would include most common summary metrics
 - Simple pie chart showing which categories came up the most
 - Table showing top requesters, and top categories per top requesters
 - Table showing top categories, and top requesters per top category
- **View Previously Categorized Cases (Top Priority)**
 - Select one or more previously categorized files to export in a single csv file
 - Export quick analysis of cases in pdf file
- **Analytics / Sorting Capabilities (Time Permitting — If we don't get to this, most of this can be done fairly quickly in Excel once the csv file is output)**
 - User would be able to select how they would like the output file to be formatted
 - Able to sort or group by year
 - Able to sort or group by customer name
 - Able to sort or group by category
 - User would be able to select which charts / tables should be included / generated
 - Pivot table, to visualize trends over time
 - Line Graph, Bar Graph, Pie Chart

Developer Requirements

The developer side of the application will maintain all of the functionality of the user requirements section. In addition to these features, the developer will have the option to retrain the neural net. This will allow for diversity between users of the application. The weights are saved when the network has been trained multiple different times, and can be switched back and forth.

Use Cases

Name	Train Neural Network
Actors	Developers
Triggers	Select “Retrain Neural Net” option from menu in developer environment.
Events	<ul style="list-style-type: none">• Select “Retrain Neural Net” button• Browse for training set .csv file• Browse for testing set .csv file• Wait for training process to run (multiple epochs)• Save new neural net weights (automatic)• Click button to go back to main menu
Exit Conditions	Neural net finishes training, or “Cancel” button is pressed (this leaves weights as they were before retraining started).
Post Conditions	New weights for the neural net are saved.
Acceptance Test	The neural net has updated weights based on the training set.

Name	Switch To Different Trained Network
Actors	Developers (possibly users later on)
Triggers	Select “Retrain Neural Net” option from menu in developer environment.
Events	<ul style="list-style-type: none">• Select “Retrain Neural Net” button• Rather than browsing for a training set, select an existing training option from the side menu• Save new neural net weights (automatic)• Click button to go back to the main menu
Exit Conditions	New weight set selected, or same weight set selected (no change).
Post Conditions	New weights for neural net are saved.
Acceptance Test	The neural net has updated weights based on the training set.

Name	Categorize New Data Set of Cases
Actors	User
Triggers	Select “New Categorization” option from menu
Events	<ul style="list-style-type: none"> • Select “New Categorization” button from menu • Browse for new data set .csv file • Wait for neural net to run • Save categorized data set (automatic) • Go to output options menu
Exit Conditions	Neural net finishes running, “Next” button is clicked, or the “Back” button is pressed to get back to the main menu.
Post Conditions	The cases in the imported .csv files are categorized.
Acceptance Test	There is a data structure holding all of the cases from the .csv import, and all categorized previously as “Other” have categories.

Name	View Categorized Data Sets
Actors	User
Triggers	Select “Already Categorized” option from menu
Events	<ul style="list-style-type: none"> • Select “Already Categorized” button from menu • View list of categorized data sets
Exit Conditions	User selects a data set or presses the “Back” button to get back to the main menu.
Post Conditions	A list of already categorized data sets is presented.
Acceptance Test	Every data set that has been categorized shows up in the list.

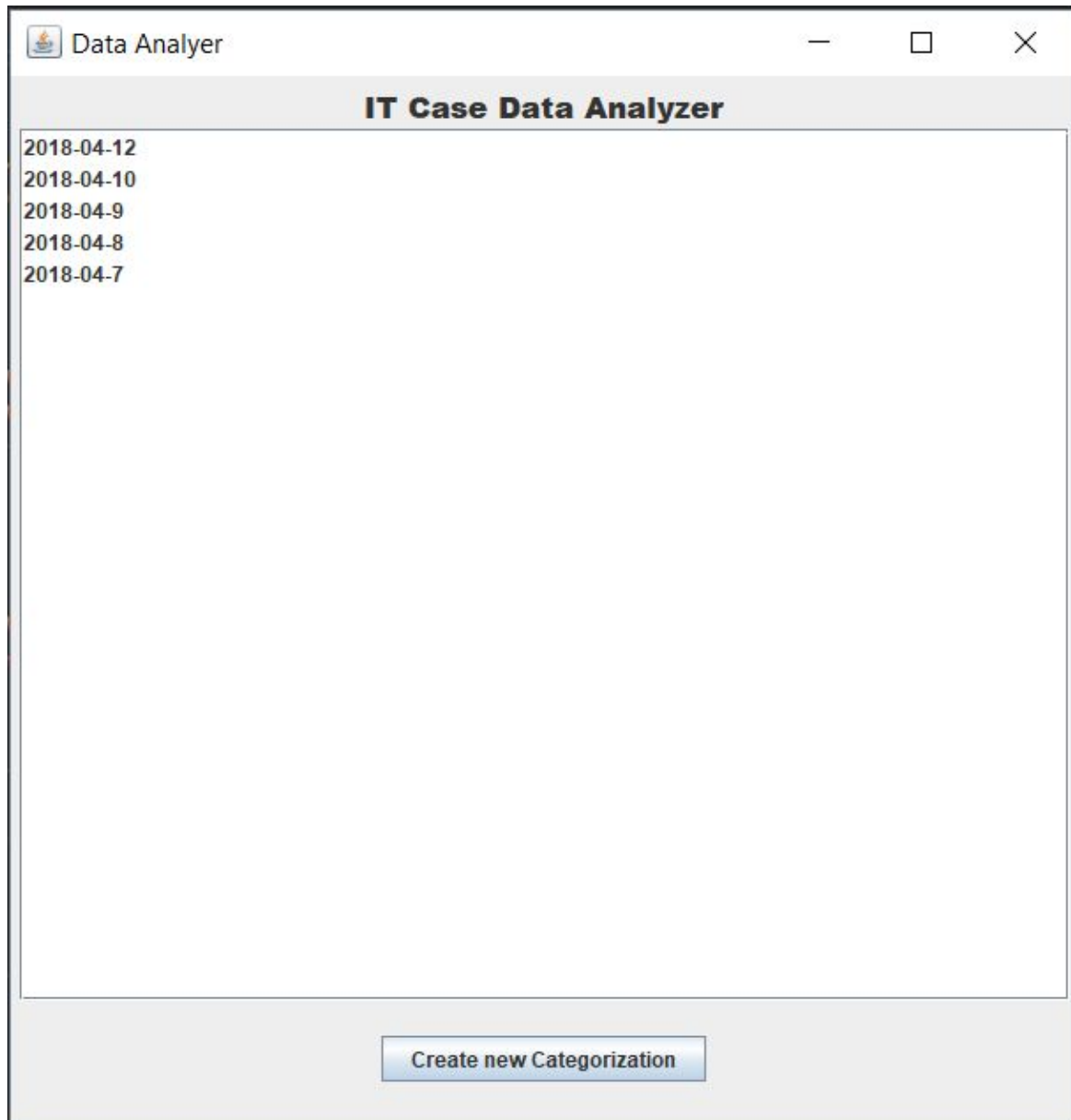
Name	Output .csv File of Categorized Data Set
Actors	User
Triggers	Select “Output .csv” from menu after new categorization, or after selecting an already categorized data set from the list.
Events	<ul style="list-style-type: none"> • Select “Output .csv” button from one of the menus described above • Wait for program to write the data to a .csv file • Select where to save the .csv file on your computer • Press the “Done” button to go back to the previous menu
Exit Conditions	“Done” button clicked or “Back” button clicked instead of saving it.
Post Conditions	A new .csv file is saved to the user’s computer with the data set fully categorized.
Acceptance Test	The .csv file can be opened in excel or similar program and correctly lists the data.

Name	Output .pdf File of Simple Metrics
Actors	User
Triggers	Select “Output Metrics .pdf” form menu after new categorization, or after selecting an already categorized data set form the list.
Events	<ul style="list-style-type: none"> • Select “Output Metrics .pdf” button from one of the menus described above • Wait for program to write the data to a .pdf file • Select where to save the .pdf file on your computer • Press the “Done” button to go back to the previous menu
Exit Conditions	“Done” button clicked or “Back” button clicked instead of saving it.
Post Conditions	A new .pdf file is saved to the user’s computer with some metrics and graphs/charts of the selected data set.
Acceptance Test	The .pdf file can be opened and contains the correct data in a visually appealing way.

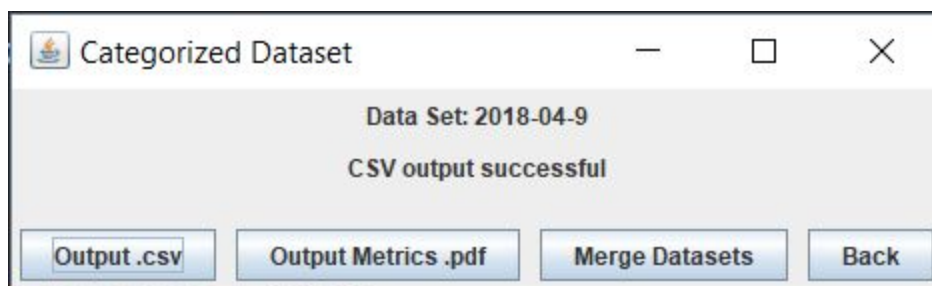
Name	Merge Datasets
Actors	User
Triggers	Select “Merge Datasets” option from menu inside “Already Categorized”
Events	<ul style="list-style-type: none"> • Select “Already Categorized” button from the main menu • Select the “Merge Datasets” button from this menu • Select which 2 datasets to merge • Wait for datasets to merge • See new dataset added which is the 2 selected ones combined
Exit Conditions	“Cancel” button clicked or “Merge” button clicked after selecting which datasets to merge.
Post Conditions	A new dataset is added to the list which contains all of the cases from each of the 2 merged datasets.
Acceptance Test	The new dataset functions fine with both output options, and contains the data from both datasets put into it.

User Interface Requirements

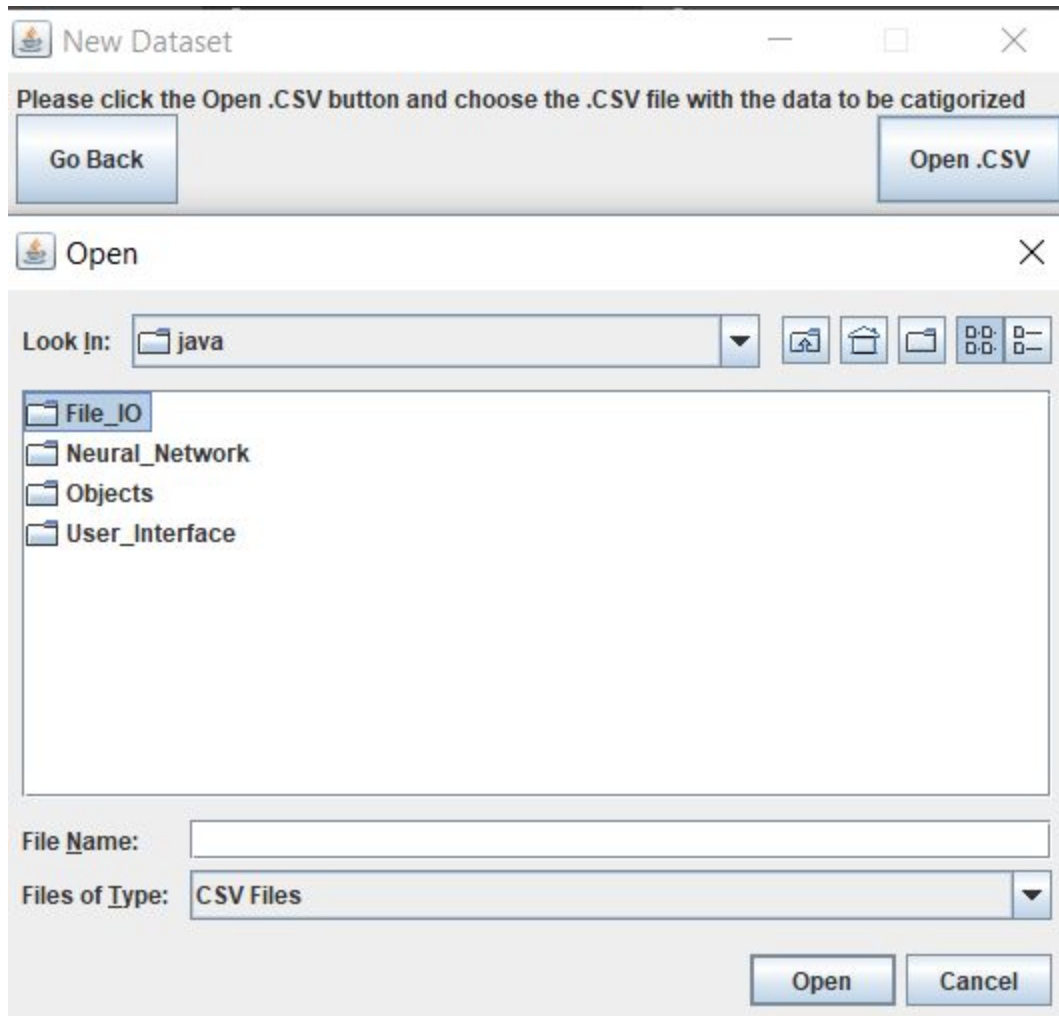
The GUI of this application will be very simple. It will open to a home screen of two options, “New Categorization” and “Already Categorized”. The former will run the neural net in order to classify new data, the giving options of outputting a .csv file and/or an output analysis .pdf. The latter choice will not need to use the neural net because it will be showing cases that have been previously categorized. They can then be outputted in a .csv file and/or output analysis .pdf.



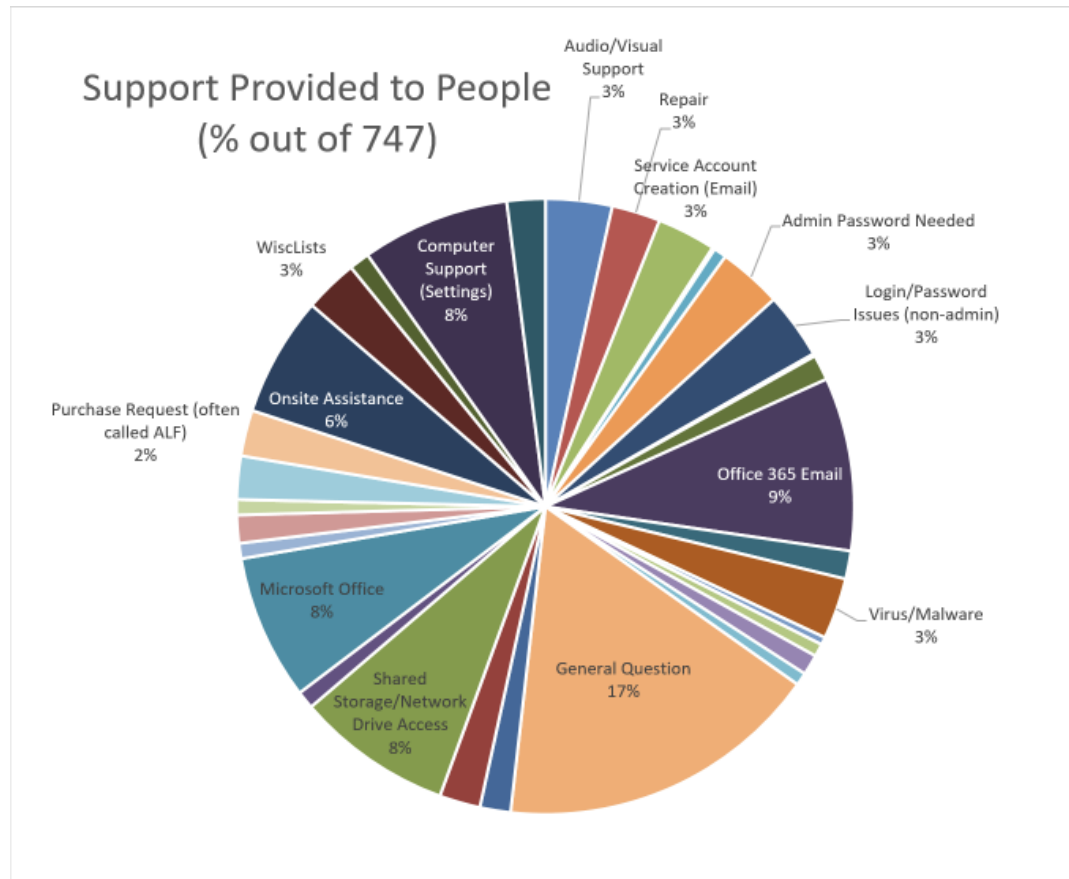
Home page



Categorized Dataset page with options



New Dataset page with Its file explorer.



Pie chart in .pdf of output analysis

Security Requirements

Sensitive data is never entered into Cherwell and therefore is not a concern for the project. If sensitive data is entered into the system, it is to be expunged from the Cherwell system by an administrator. Essentially, we have no security concerns beyond pulling data between a mistaken entry and removal by an admin. If this does occur the user can manually remove it from the data pool due to the accessibility of our long-term data storage. There is no need for encryption or logins as this software is primarily parsing data on the system and placing it back into the customer's system.

System Requirements

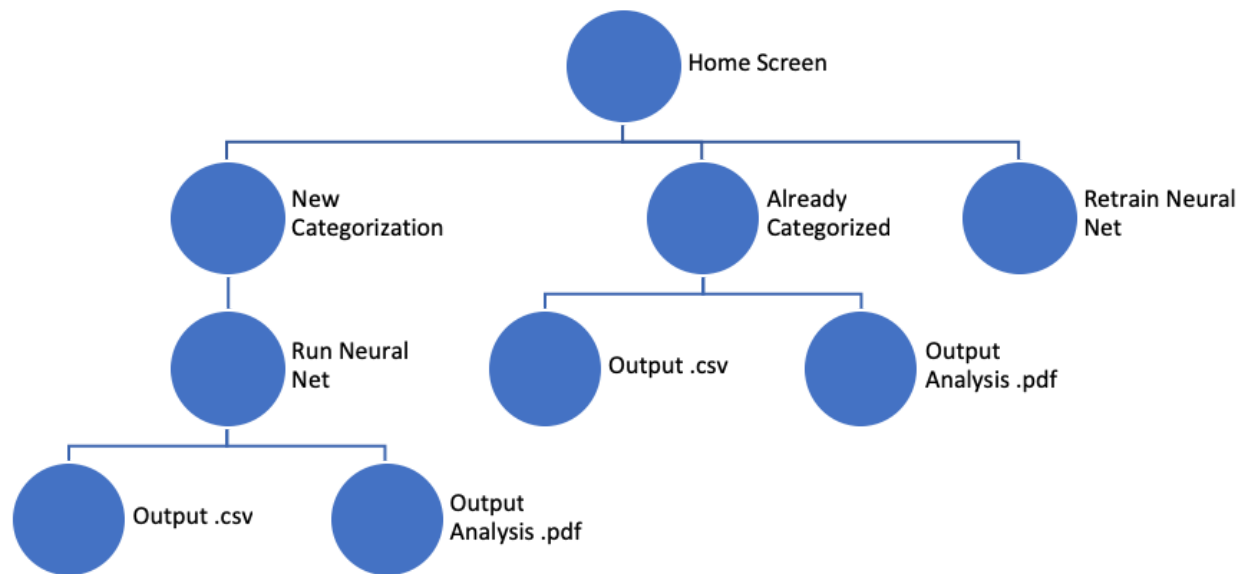
The application will only need to be compatible with Windows as that is the primary operating system used by DoIT employees and the Windows version of Excel is preferred by Patrick Johnson, the primary user. The application will not need to be compatible with a Unix based system. This application will need to be able to interact with data in the CSV format and be able to input data from and output data to Microsoft Excel. Additionally, this application needs to be capable of outputting to a PDF file. The average CSV file size that the application will be

inputting is around 120 KB. The CSV file that will be outputted will be approximately the same size, and the PDF output will be slightly larger. Since there may be a “merge files” option in the application there could be an output file around 300 KB. As a result, the application will need to be capable of processing and storing files around 150 KB on average with a maximum of around 500 KB.

The user will need Java installed on their device to run the application, and to get good use of the output, excel and a PDF reader, though the PDF could also be viewed in any browser if they do not have access to a dedicated program.

Specification

Full Application Layout

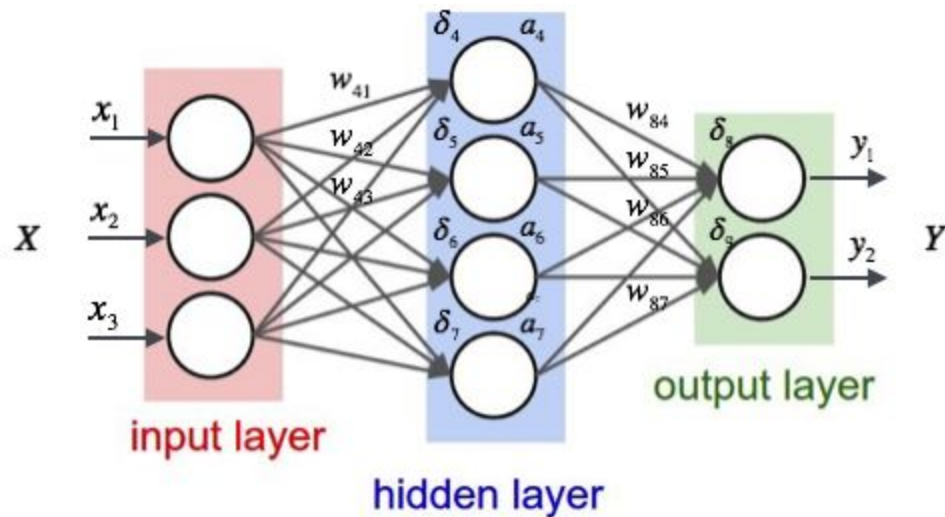


The application is relatively simple in terms of the functions it will offer. When starting it up there are just 2 options to choose from, which are running a new data set through the neural net to categorize them, or view previous data sets that have already been categorized. Once a data set of cases is categorized there are 2 planned output options

- .csv file with a pivot table to make analyzing the data easy in Excel or Tableau
- .pdf file with most common metrics explained, and simple Pie Chart showing which category came up the most
 - Top Requesters
 - Top Categories per top requesters
 - Top Categories
 - Top Requesters per top categories

For now the “Retrain Neural Net” option would only be accessible in our development environment, not to the user. It would take it a training set and a testing set, use gradient descent to train the network, and then figure out how accurate it is against the testing set. When the network has been trained multiple different times, the weights are saved and can be switched back and forth between.

Neural Net Concept

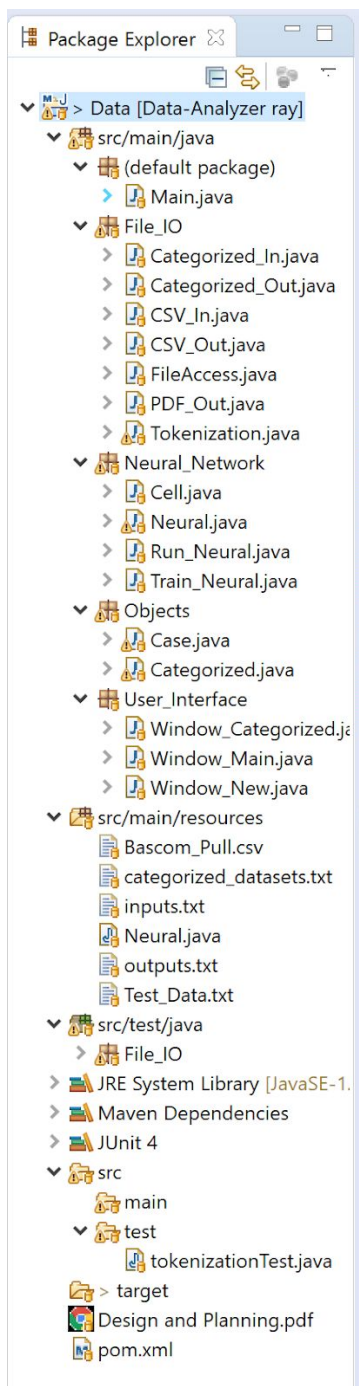


The neural net design is the most important part of the application, and will determine how accurate categorization is. Inputs will be the counts of certain keywords in a case description, and outputs will be the categorization.

- Inputs (We may end up using fewer inputs in future iterations if training takes too long)
 - We have 153 keywords that are used as inputs
 - Examples include “office”, “365”, “firewall”, “projector”, “scam”, etc.
- Output Categories (These categories are likely final, but in future iterations we plan on reordering them so similar categories are next to each other)
 - Office 365 Email and Calendar
 - Shared/Network Drive
 - Microsoft Office
 - Computer Support Settings
 - Purchase Request
 - Onsite Assistance
 - Wisclists
 - Audio/Visual Support
 - Device Repair
 - Service Account Creation (Email)
 - Admin Account/Password
 - Login Issues
 - Virus/Malware
 - Printer Support
 - Multi Factor Authentication
 - Room Access
 - Network Connectivity
 - Adobe Suite
 - Loaner Request (Computer/Mifi)

In terms of verifying the performance or accuracy of our neural network, we can run our trained network on the test data (or any set of cases that have been manually categorized), and then count how many cases it correctly categorized, how many it incorrectly categorized, and how many it simply placed in “General Question”. We may develop a more sophisticated measure in future iterations (once similar categories are next to each other), but for now our measure is planned to be as follows: Each correctly categorized case will count as +1, each incorrectly categorized case will count as -1, and a case placed in “General Question” will count as 0. Then, sum these up and divide by the number of cases. Possible values will range from -1 to 1, with 1 implying perfect categorization and -1 implying entirely miscategorized.

Project Folder Layout



“WiscIT” Spitout Example .csv (Input to our Program)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Incident Type	Incident ID	Customer ID	Owned By	Created Date	Status	Description	SLA Resolve	Last Modified	Specific Type	Call Source	Repair Cover	Category	Subcategory		
2	Incident	1000319	KATIE L P BLI	David Davenj	8/4/14 9:09	Closed	Katie is tryin	8/9/14 9:09	Christopher Grosspietsch	Phone			Interactive R	Technical Issue/Site		
3	Incident	1000334	JOSHUA MO	Latif Lewis	8/4/14 9:28	Closed	Customer is	8/9/14 9:28	Christopher Grosspietsch	Phone			WISCMail PI	Submit Incident		
4	Incident	1000374	CHRISTOPHE	William Crid	#####	Closed	Customer ne	#####	Christopher (Handling Spe	Phone			Endpoint Ma	Submit Incident		
5	Service Reque	1000411	TODD D SCH	Mason Hoeft	#####	Closed	Customer is	#####	Christopher Grosspietsch	Portal			CAMPUS NE	Request/Data Jack/Activation		
6	Incident	1000555	JOSHUA MO	Jason Power	#####	Closed	Student life i	#####	Christopher Grosspietsch	Phone			STUDENT CE	Submit Incident		
7	Incident	1000581	TODD D SCH	Jon Dickenso	#####	Closed	Customer is	#####	Christopher Grosspietsch	Phone			WISCMail PI	Submit Incident		
8	Incident	1000588	JOHN E WAL	Jason Power	#####	Closed	Customer is	#####	Christopher Grosspietsch	Phone			PERSONAL S	Personal Software/Email		
9	Incident	1000673	JAYNE N NT	William Crid	#####	Closed	Norah report	#####	Christopher (Handling Spe	Portal			Endpoint Ma	Submit Incident		
10	Incident	1000695	GEOFFREY V	Jon Dickenso	#####	Closed	Geoff has re	#####	Christopher (Handling Spe	Portal			Endpoint Ma	Submit Incident		
11	Incident	1000844	KRISTOPHER	Nicholas Anc	8/5/14 9:10	Closed	Customer wi	#####	Christopher Grosspietsch	Portal			eduroam	Submit Incident		
12	Incident	1000889	SHELTEREESE	Jon Dickenso	8/5/14 9:55	Closed	Needs to hav	#####	Christopher Grosspietsch	Phone			PERSONAL S	Personal Software/Email		
13	Incident	1001255	TODD D SCH	William Crid	#####	Closed	Todd reports	#####	Christopher (Handling Spe	Phone			Endpoint Ma	Submit Incident		
14	Incident	1001296	ESMERALDA	Frederic Mur	#####	Closed	Esmeralda w	#####	Christopher (Handling Spe	Phone			Endpoint Ma	Submit Incident		
15	Incident	1001395	KATHRYN M	Marc Mrozin	8/6/14 8:36	Closed	Customer is	#####	Christopher Grosspietsch	Phone			Campus Net	Submit Incident		
16	Incident	1001421	LAURIE LEIN	Rhianna Carr	8/6/14 9:18	Closed	User reporte	#####	Christopher Grosspietsch	Email			WISCMail PI	Submit Incident		
17	Incident	1001433	KORY R DEAN	Jon Dickenso	8/6/14 9:38	Closed	User request	#####	Christopher (Handling Spe	Email			Endpoint Ma	Submit Incident		
18	Incident	1001519	CAROL J POP	Thomas Reir	#####	Closed	8/15/14 Dis	#####	Christopher Grosspietsch	MySoft Order			VOICE SERVI	Submit Incident		
19	Repair	1001596	TODD D SCH	Jacob Lewis	#####	Closed	User reporte	#####	Christopher (Handling Spe	Email			Device Repai	Submit Repair Request		
20	Incident	1001667	KRISTINE K F	Jon Dickenso	#####	Closed	There is a wi	#####	Christopher Grosspietsch	Phone			General Dep	Wireless Keyboard		
21	Incident	1001691	DONNA M E	Thomas Reir	#####	Closed	8/12/14\r\nl	#####	Christopher Grosspietsch	MySoft Order			VOICE SERVI	Submit Incident		
22	Incident	1001733	TRACY JOY N	Andrew Aubi	#####	Closed	After people	#####	Christopher Grosspietsch	Phone			QUALTRICS S	Submit Incident		
23	Repair	1001751	MEGAN E M	Edward Your	#####	Closed	Megan work	#####	Christopher (Handling Spe	Phone			Device Repai	Submit Repair Request		
24	Incident	1001752	MEGAN E M	Jon Dickenso	#####	Closed	Megan has r	#####	Christopher (Handling Spe	Portal			Endpoint Ma	Submit Incident		
25	Incident	1001767	TODD D SCH	Latif Lewis	#####	Closed	Customer is	#####	Christopher Grosspietsch	Email			WISCMail	Submit Incident		
26	Incident	1001775	VALERIA A D	Stefanie Des	#####	Closed	Customer is	#####	Christopher Grosspietsch	Phone			SFS - E-REIV	Submit Incident		
27	Incident	1001861	TODD D SCH	Marc Mrozin	8/7/14 7:28	Closed	Customer is	#####	Christopher Grosspietsch	Phone			WISCMail PI	Submit Incident		
28	Incident	1001870	MICHAELA A	Jon Dickenso	8/7/14 7:39	Closed	Customer ca	8/9/14 7:39	Christopher Grosspietsch	Phone			MAPPING DF	Submit Incident		
29	Incident	1001887	REX GERARD	Jon Dickenso	8/7/14 8:16	Closed	There was a	#####	Christopher Grosspietsch	Phone			General Dep	Shared Drive Access Request		
30	Incident	1001999	NICOLE A BR	Andrew Aubi	#####	Closed	Customer ne	#####	Christopher Grosspietsch	Phone			PERSONAL S	Personal Software/Chat		
31	Incident	1002155	DIANE L STEI	Marc Mrozin	#####	Closed	Diane is won	#####	Christopher Grosspietsch	Phone			WISCMail	Submit Incident		
32	Incident	1002176	LINDSEY N D	Andrew Aubi	#####	Closed	Wants to mc	#####	Christopher Grosspietsch	Phone			General Dep	Software Install		
33	Incident	1002198	DARREN J U	Joseph Leroy	#####	Closed	Customer ca	#####	Christopher Grosspietsch	Phone			WISCMail	Feedback		
34	Incident	1002209	KATIE L P BLI	Christopher I	#####	Closed	The custome	#####	Christopher Grosspietsch	Phone			SIS	Submit Incident		
35	Incident	1002267	TODD D SCH	Jon Dickenso	#####	Closed	Mail Messag	#####	Cherwell Admin	Email			WiscIT	Unhandled Inbound Email		
36	Service Reque	1002297	JESSELAN EN	Gary Norther	#####	Closed	1254 - Basco	#####	Christopher (Handling Spe	Phone			CAMPUS NE	Request/Data Jack/Installation		
37	Repair	1002495	DAVID M BE	Caitlin Staley	8/8/14 9:56	Closed	Customer ca	8/8/14 9:56	Christopher (Specifics - Di	Walk-In (Con Time&Mater			Device Repai	Submit Repair Request		
38	Incident	1002512	GEOFFREY V	Jason Power	#####	Closed	Customer ca	#####	Christopher Grosspietsch	Phone			WISCMail	Submit Incident		
39	Incident	1002545	JOHN E WAL	Jason Power	#####	Closed	Customer ca	#####	Christopher Grosspietsch	Phone			WISCMail	Submit Incident		