

**Helpdesk Support Ticket   
Auto-Classifier Model**

*July 2022*

**Derek Powell**

**Business Problem:**

In the world of technical support, identifying patterns and trends within the issues reported by users is critical for understanding solutioning gaps and making a plan towards addressing them. While some patterns and trends are easy to see, especially at a micro level, it can become problematic to identify quality patterns at a macro level.

For this reason, most ticketing (or CRM) software used by support desks includes drop-down selection fields for agents to categorize the nature of the issue within the ticket. However, this presents a set of key challenges:

* As the complexity of the solution grows, there arises more and more categories of interest, and there comes more and more hierarchical selections, and the complexity of simply categorizing the issue grows.
* As the complexity of categorizing the issue grows, so too does the cost of training help desk agents as well as likelihood of incorrect selections.

For large companies with complex solutions, categorizing every ticket accurately presents a very high training cost and even still then a high probability of miscategorization.

This project seeks to create a machine learning model which can evaluate features within a help desk ticket, including user commentary, and automatically (and accurately) classify the issue. This would greatly improve the value of this data in the hands of Product, Development, and Support organizations and would greatly reduce training costs.

**Datasets:**

Within this project, I will be leveraging data found within my company’s ticketing software, seeking to take a recent, and well-rounded, sample of cases. The data is laden with features, representing each field found within a ticket, including discrete variables, continuous variables, and date/time objects.

I expect to find everything I need within the same data warehouse and data export.

**Anticipated Data Science Approach:**

The data available has already been categorized using a hierarchical series of categorization selections made by agents, however the options available are limited in some cases (do not offer the business value desired, even when accurately selected) and even where options are not limited, selections made by agents are inconsistent and lack accuracy, making the data difficult to trust.

For that reason, I plan to approach the data as ‘unlabeled’ and use an unsupervised classification model for initial hierarchical segmentation. Once initial segmentation has been done, I plan to manually evaluate each observation and fine-tune the classification by hand.

Once an acceptable amount of data has been accurately labeled using unsupervised classification and hand-tuning, I will use that data to train a supervised classification model to accurately predict classification of unseen data.

**Deliverables:**

* The completed code, including the model and inputs for entering new data as needed.
* A project report, including detailed instructions how the model was trained and how to approach retraining, or improving of the model, in the future if needed.