**Escalated Call Prediction**

**Business Problem**

Contact centers are designed to handle customer inquiries and issues in an efficient manner. Time is money so every second on a call may be a cost to the company. Of course, servicing customers in a positive way can impact business and produce benefits also. The goal of this project is to detect if calls are starting to show signs of needing to be transferred to another agent more skilled to handle escalated situations or highly complicated issues. Routing calls with real potential for escalation is critical to maintain customer service quality and mediate issues before they advance.

**Background/History**

I will search for words in the call transcripts that may indicate escalation such as when a customer is asking for a manager or supervisor to label the records in the data with a binary indicator. I will then expand on that concept but rather than reviewing data after it occurs, we can predict the escalation probability of a call in closer to real-time. I will use transformers to find the semantic meaning behind sentences within call transcripts and then use a neural network classifier to estimate the potential for escalation.

**Data Explanation**

I used Hugging Face to retrieve a group of files containing real-world call transcript data from contact centers. This data was collected from several different industries globally and translated into English. There are over 91,000 transcripts available for analysis. I downloaded each of the folders containing JSON files to the same directory as my Jupyter Notebook to get the data to load. I had to iterate through the files to extract only valid files and records since there were several JSON formatting errors in the files. I refined the escalation patterns to search in the transcripts. I labeled the records in the data to indicate if they match the escalation patterns that I established (such as asking for a manager or supervisor). I will also use Natural Language Processing to reduce words to root word and remove punctuation and stop words.

**Methods**

After pre-processing the data, I loaded 20,000 of the clean records of the transcripts with the escalation label indicator. I used SentenceTransformer to interpret words within the context of the sentence and sentences within the context of the transcript. The transcripts are then encoded based on their general meanings. Since there was a class imbalance between typical calls and escalation potential calls, I oversampled from the escalation calls using SMOTE (See Figure 1).

Figure 1: Class Distribution before and after oversampling minority class

A comparison of blue squares

AI-generated content may be incorrect.

You can see the imbalance of negative words that may indicate a potential for call escalation in Figure 2.

Figure 2: Wordcloud of transcripts (shows mostly neutral and positive sentiments)A close-up of words

AI-generated content may be incorrect.

I used a Multi-Layer Perceptron learning algorithm (Scikit-learn developers, n.d.) to create a neural network that could use the transformed sentences and avoid overfitting to the model to use to predict the escalation probability of a record. I used a neural network model since I was working with a large dataset containing language that needed to be classified based on intent of call and whether it indicated an escalation potential. I used 30 epochs to train the 70% of the split dataset on the neural network classifier.

**Analysis**

I will use metrics such as balancing F1 score and a high recall of escalated calls to see how well the model performs and if it needs to be tuned more. Will review varying probability thresholds to find the best one to use for maximizing recall in the escalation prediction in the model. Figure 3 shows how current tuning is performing when compared to perfect prediction.

Figure 3: Actual versus ideal predictions

A graph of a graph

AI-generated content may be incorrect.

**Conclusion**

This data pipeline will be able to perform well in predicting potential escalation of calls. It will be a great benefit to routing calls accurately. Agents skilled to handle escalation calls can be ready and available when needed to handle call escalations.

**Assumptions**

I had datasets from a variety of service areas, automotive and home insurance, to name a couple of the areas. An assumption with this project is that the transcripts would represent the conversation and tone and that the model would be able to generalize well across different areas.

**Limitations**

A limitation that is inherent in this pipeline is the reliance on escalation patterns in transcripts to label the data. The keywords and patterns being used may differ across the records from the JSON files and may not get captured. These will need monitored and changed as new transcripts are available.

**Challenges/Issues**

The loading of the datasets presented a challenge, but after much trial and error I was able to sort through and then load enough usable transcripts for analysis. There was also a class imbalance to overcome with the original sample of data, so I resampled it with a focus on oversampling from the minority class (potential for escalation). The biggest issue I see for future maintenance of the model will be deciding which keywords to use for escalation and how often to fine tune the parameters of the model.

**Future Uses/Additional Applications**

This data pipeline could be expanded to use for real-time routing of calls and categorizing calls to be directed to agents specialized in escalation for the customer’s issue. We could also solicit feedback from the agent to see if a call should have been flagged for escalation but was not or feedback from the higher skilled agent for calls that did not need escalated. This could also be used for identifying issues when escalations for certain categories begin to rise or could be applied to messaging or chat interactions rather than only calls.

**Recommendations**

One recommendation is to add real escalation records to the training data after the initial model is implemented. This will improve the learning of the neural network classifier by having better records to learn from.

**Implementation Plan**

Present to call center directors or those in charge of telephone call routing. Show how model can predict potential for escalation based on the labeling and training the data. Once escalation classifier model is in use in their contact center, it will trigger an indicator to route calls with potential for escalation to more uniquely skilled agents. Based on labels (feedback) provided by agents, those call records will go into training the model to keep it performing well.

**Ethical Considerations/Assessment**

The dataset has carefully redacted the personally identifiable information (PII) from the transcripts. This eliminates the risk of non-compliance to data privacy laws in using this data. The transcripts should be monitored or reviewed for new keywords or transcript sentiment to ensure that model is still being trained to correctly classify the records. Care also needs to be taken when adding keywords to the escalation patterns to ensure that they are bias to any specific demographics.

**Audience Questions/Answers**

1. How accurate is the model at identifying escalated calls?
   * The model achieves a high recall for the escalated calls, meaning that it is correctly identifying most of them. This is critical for routing the sensitive calls to the higher skilled agents.
2. How does the model know if the call is a potential for escalation?
   * It uses patterns in the data to predict potential escalation. The patterns are learned from thousands of transcripts and can be refined based on feedback, if implemented.
3. What’s the benefit to this system compared to our current call routing process?
   * Proactive instead of reactive identification of escalating calls. Reduced resolution times, customer satisfaction improvements, and reduced risk from complaints.
4. Can this model reduce costs?
   * This model can help reduce call handling times which increases operational efficiency and staffing expenses.
5. What happens if the model makes a wrong prediction?
   * We are proposing to set up a feedback loop from agents where their feedback can go into training the model to further refine the process and reduce false positives.
6. How does the system handle that escalated calls are not common?
   * I used SMOTE to oversample from the escalated calls population of the training data to have a more balanced dataset to train the model.
7. Does this model understand tone or strictly input keywords/patterns?
   * Yes, the use of a natural language model can recognize context and meaning to identify escalation even if expressed with different keywords.
8. Can this be adapted to different call types or functions?
   * Definitely. This model was created using transcripts from varying call center domains.
9. What kind of support will this system require?
   * Periodic monitoring of model performance and possible hyper tuning as new call records and feedback is added to training model.
10. When are you available to discuss implementing this model for our call center?
    * Immediately. Thank you for your interest.

**References**

AlxBlock. (2025). *92k-real-world-call-center-scripts (English)* [Data set]. Hugging Face. <https://huggingface.co/datasets/AIxBlock/92k-real-world-call-center-scripts-english>

OpenAI. (2025). ChatGPT (October 25 version) [Large language model]. <https://chat.openai.com/>

Scikit-learn developers. (n.d.). 1*.17. Neural network models (supervised).* In *scikit-learn: Machine Learning in Python* (Version 1.7.2). Retrieved October 18, 2025, from <https://scikit-learn.org/stable/modules/neural_networks_supervised.html>