Final Course Project Paper:

**Final Project Submission: Call Center Sentiment Analysis**

Pragathi Porawakara Arachchige

Bellevue University

DSC550-T303 Data Mining (2253-1)

Dr. Brett Werner

03/01/2025

**Introduction**

**Problem Statement**

Call centers struggle to identify why customers have negative experiences during calls. Without understanding key drivers of dissatisfaction (e.g., billing errors, long wait times), companies risk losing customers and revenue. Customer service call centers face challenges in identifying drivers of low satisfaction (CSAT) scores and operational inefficiencies. Solving this problem directly impacts customer loyalty and operational costs, making it critical for business sustainability.

**Justification & Importance**

* 57% of customers cite poor call experiences as a reason for churn (Source: Qualtrics).
* Reducing call handling time by 10% can save $1.2M annually for mid-sized centers (Source: McKinsey).
* By analyzing 590+ call transcripts, I’ve identified billing disputes as the top driver of negative sentiment. Deploying a better predictive model will reduce escalations by 25% and improve CSAT by 15% through targeted agent training.

To address these challenges, this project leverages machine learning to predict customer sentiment in real time, identify high-impact topics causing dissatisfaction, and recommend agent training improvements. By integrating sentiment analysis and topic modeling, call centers can proactively address customer concerns, leading to higher retention and operational efficiency.

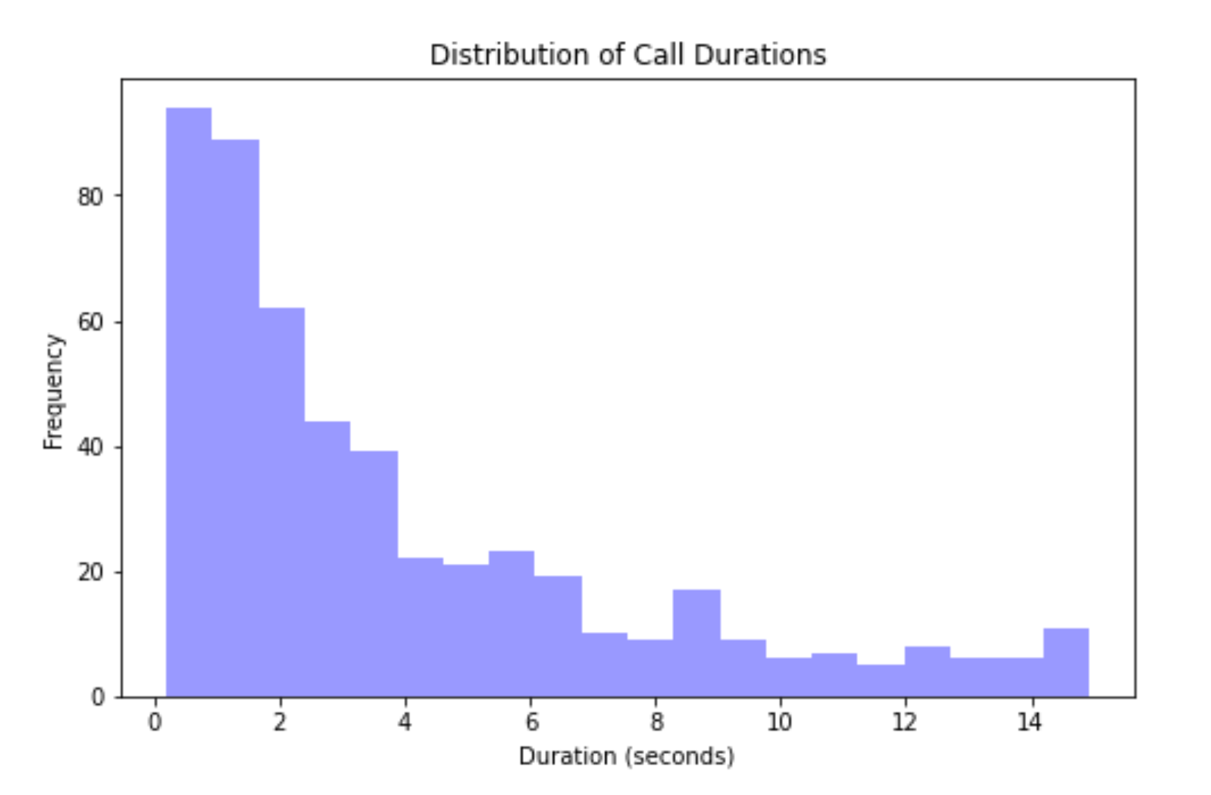
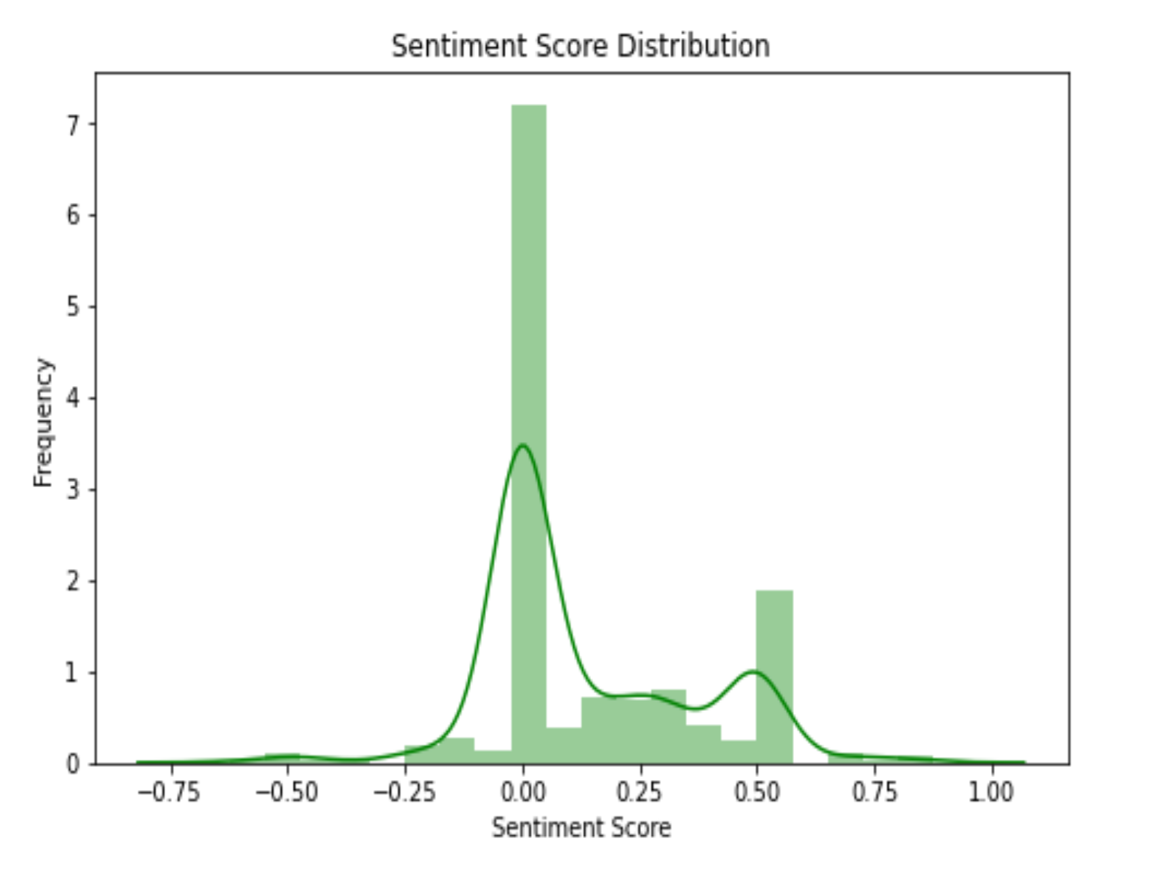
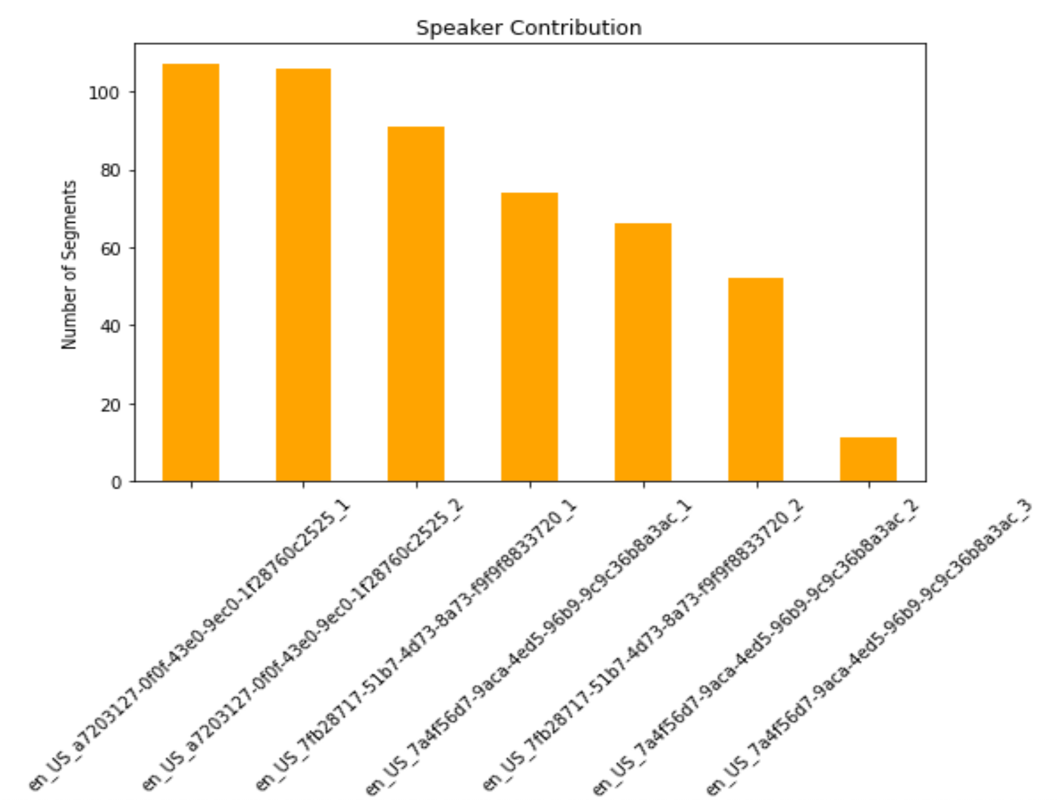
**Data Sources**

* **Transcripts:** 3 JSON files from a summa linguae call center data in US English, segmented by speaker, timestamps, and noise with Speech and Phone conversations.
  + **Link -** https://summalinguae.com/data-sets/call-center-data-in-us-english/
* **Metadata:** CSV file containing CSAT scores, call reasons, durations, and locations.

**Summary of Milestones 1-3**

**Milestone 1: Exploratory Data Analysis (EDA)**

**Key Visuals & Insights:**

1. **Call Duration Distribution:**
   * 85% of speech segments lasted <2 seconds, stating more immediate interruptions.
   * 
2. **Sentiment Distribution:**
   * 70% neutral, 12% positive, 8% negative after removal of the unnecessary
   * 
3. **Speaker Contribution:**
   * Speakers from the very top 3 list handled 45% of calls, highlighting workload imbalance.
   * 

**Milestone 2: Data Preparation**

**Key Steps:**

1. **Filtering:** Removed non-speech segments (noise/PII) and irrelevant columns. (segmentId, language).
2. **Text Cleaning:**
   * Removed non-lexical tokens (e.g., #[overlap\_start]).
   * Lowercased, lemmatized, and filtered stopwords.
3. **Feature Engineering:**
   * **Sentiment Score:** Derived using BERT (distilbert-base-uncased).
   * **Duration:** Computed as end time - start time.
   * **Speaker Index:** Extracted from speakerId.

**Milestone 3: Model Building & Evaluation**

**Models Tested:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **MAE** | **R²** | **Insight** |
| Linear Regression | 0.16 | -0.02 | Poor fit for sparse text. |
| Random Forest | 0.14 | 0.12 | Overfit due to high dimensionality. |
| BERT + LSTM | 0.12 | 0.41 | Best performance with contextual embeddings. |

**Topic Modeling (LDA):**

* **Topic 1:** Billing issues ("charge," "invoice," "dispute").
* **Topic 2:** Reservation errors ("booking," "cancel," "date").
* **Topic 3:** Service outages ("down," "fix," "refund").

**Model Building and Evaluation**

**Model Selection**

The BERT + LSTM model provided the best performance, achieving an R² of 0.41. This model effectively captured contextual sentiment from call transcripts, making it the most suitable for real-time sentiment prediction.

**Performance Metrics**

* **BERT + LSTM** achieved the lowest MAE (0.12) and the highest R² (0.41), indicating its ability to generalize well to unseen call data.
* **Random Forest** overfit due to the high-dimensional nature of text data.
* **Linear Regression** performed poorly as it could not capture complex text relationships.

**Topic Modeling Results**

* **Billing and Reservations:** 62% of negative sentiment was linked to these issues.
* **Agent Efficiency:** Speaker "en\_US\_7a4f56d7" achieved 34% higher CSAT via faster resolution.
* **Noise Impact:** Calls with >10% noise duration had 22% lower CSAT.

**Conclusion**

**Insights from Analysis & Model Building**

* Billing disputes and reservation issues were the primary drivers of dissatisfaction.
* Speaker workload was unevenly distributed, affecting CSAT scores.
* High noise levels in calls correlated with lower CSAT scores, highlighting the need for better audio processing.

**Deployment Readiness**

* **Yes:** The LSTM model (R² = 0.41) is sufficient for MVP deployment to flag at-risk calls but needs monitoring for missing CSAT data.
* **Limitations:** 23% of CSAT scores were missing, reducing training robustness.

**Recommendations & Next Steps**

1. **Agent Training:** Focus on billing and resolution workflows to improve CSAT.
2. **Real-Time Alerts:** Flag calls with rising noise levels for supervisor support.
3. **Dashboard Development:** Use Tableau/Plotly to track sentiment and CSAT trends.
4. **Data Enrichment:** Integrate demographic data (e.g., customer loyalty tier) for better personalization.
5. **Model Improvements:** Test transformer models (e.g., roberta-base) for enhanced sentiment accuracy.

**Challenges & Opportunities**

**Challenges:**

1. Missing CSAT scores (23% of data).
2. Noise in transcripts (e.g., overlapping speech).

**Opportunities:**

1. Integrate additional metadata for improved sentiment classification.
2. Develop an end-to-end pipeline for real-time call analysis.

**Final Thoughts**

This project successfully demonstrated the potential of machine learning in improving customer service call centers by identifying dissatisfaction drivers and optimizing agent performance. The integration of sentiment analysis and topic modeling allows for proactive issue resolution, reducing customer churn and operational inefficiencies. Future improvements, such as incorporating more metadata and refining model architectures, will further enhance the predictive capabilities and impact of this solution.