

Chat Query Router Analysis Report

Model Comparison and Classification Results

March 29, 2025

Author: Gilles Ferrero

gilles.ferrero@gmail.com

Project's Github: <https://github.com/gife2907/NLP-BERT-Query-Router>

gilles.ferrero@gmail.com

Table of Contents

Table of Contents	2
Introduction.....	4
Work summary.....	5
Abstract: Comparative Analysis of Transformer Models for Chat Query Routing in Lawn Care Support Applications.....	5
System Diagram.....	6
12 Key ML/NLP Features in this Classification System.....	7
Python libraries and models used for this ML/NLP classification job:	7
Python Libraries:.....	7
ML/NLP Models:.....	7
Topic Classification Structure.....	8
Model: bert-base-uncased	8
Close Categories Analysis.....	8
Test Data.....	10
Classification Results	10
Model: prajjwal1/bert-mini	13
Close Categories Analysis.....	13
Test Data.....	14
Classification Results	14
Model: all-MiniLM-L6-v2	17
Close Categories Analysis.....	17
Test Data.....	17
Classification Results	17
Category-Level Model Comparison.....	20

Category+Subcategory-Level Model Comparison.....	20
Execution Summary	20
Appendix	21
Classification Glossary	21

Introduction

About Me

I am Gilles Ferrero, a senior software engineer and director of engineering specializing in Natural Language Processing (NLP), Machine Learning (ML) and Artificial Intelligence (AI) applications for business. With over 25 years of experience in the Software R&D Industry, I have developed numerous solutions for Fortune 500 companies and startups alike.



Expertise

My current R&D works focuses on developing novel approaches to text classification, sentiment analysis, and entity recognition. I hold a Master of Science in Computer Science from Telecom Nancy University in France.

Previous Work

I've successfully implemented NLP solutions across various domains including:

- B2C: Web page filtering classifiers to protect minors on the internet
- Telco: Server based content classifiers for content classification at core network levels
- ECommerce: Chat request classifier/router to support end customers
- Retail: Sentiment & Customer feedback analysis over product recommendations

My Approach

Each project begins with a thorough understanding of the business problem, followed by data exploration and prototype development. I prioritize interpretable models that provide not just accurate predictions, but actionable insights that drive business value.

The analysis presented in this report follows my methodical approach to model development, rigorously testing multiple architectures to identify the optimal solution for your specific use case.

Work summary

Abstract: Comparative Analysis of Transformer Models for Chat Query Routing in Lawn Care Support Applications

This research evaluates three transformer-based models (BERT-base-uncased, BERT-mini, and MiniLM-L6-v2) for their effectiveness in implementing a chat query router within a lawn care customer support system. The system employs a zero-shot classification approach using semantic similarity to categorize customer queries into a two-level taxonomic structure without requiring explicit training.

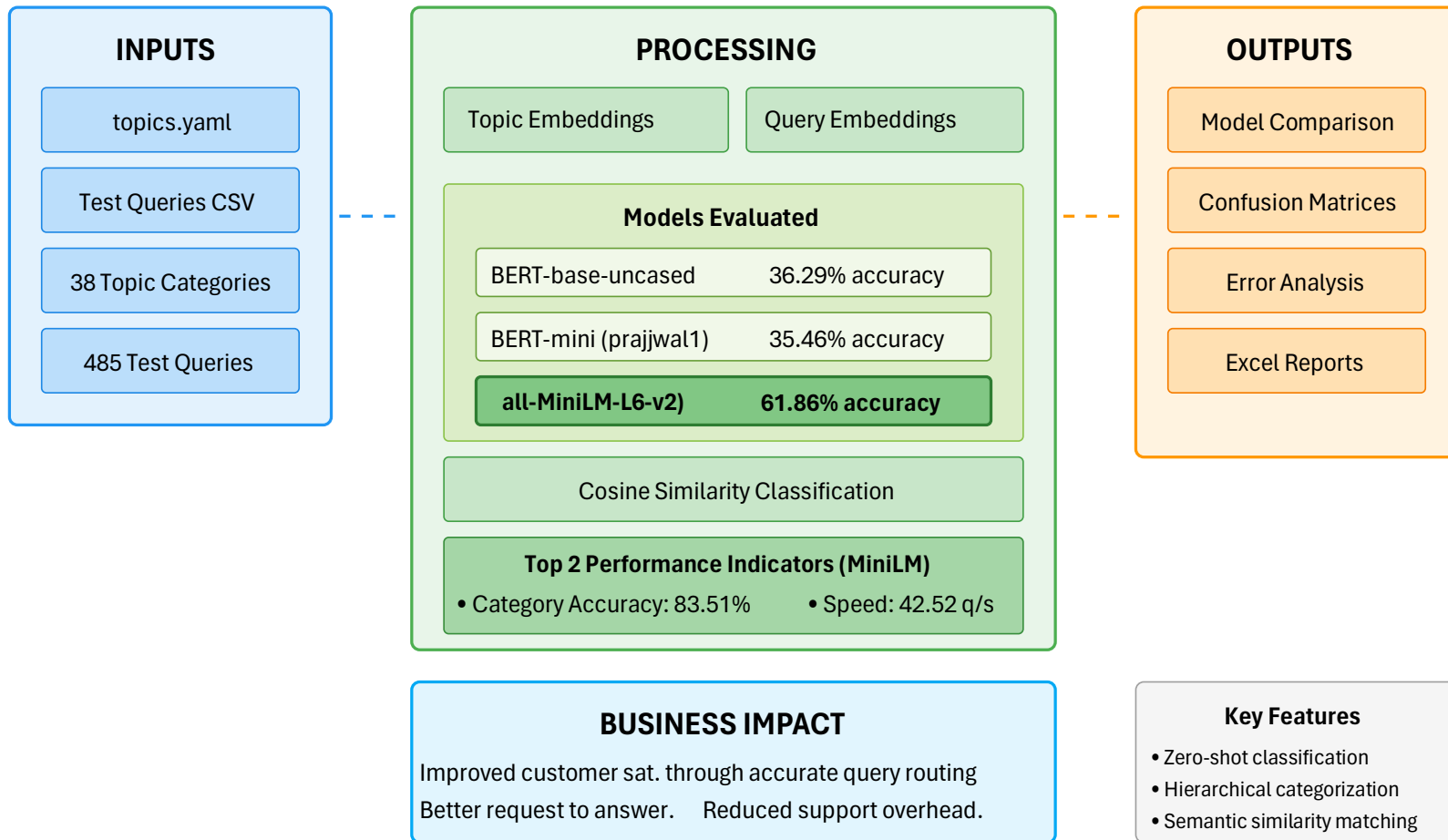
The study examined 485 lawn care related queries across 38 topic categories. Performance was assessed on both category-level accuracy and the more challenging category+subcategory level. MiniLM-L6-v2 demonstrated superior performance with 83.51% category-level accuracy and 61.86% subcategory accuracy, significantly outperforming both BERT-base-uncased (76.29%/36.29%) and BERT-mini (72.16%/35.46%).

Model	Accuracy	Precision	Recall	F1 Score	Queries/sec	Time/query
bert-base-uncased	76.29%	77.81%	76.29%	74.65%	12.22	81.81 ms
prajjwal1/bert-mini	72.16%	81.00%	72.16%	71.90%	43.87	22.80 ms
all-MiniLM-L6-v2	83.51%	83.78%	83.51%	83.22%	42.52	23.52 ms

Notably, MiniLM-L6-v2 achieved this superior performance while maintaining fast inference speeds (42.52 queries/second), comparable to the lightweight BERT-mini (43.87 queries/second) and substantially faster than BERT-base-uncased (12.22 queries/second). The analysis also identified semantically similar categories that could potentially be merged or redefined to improve classification boundaries.

This work demonstrates that compact embedding models can outperform larger alternatives in specialized domain classification tasks, providing both higher accuracy and faster processing speeds for real-time query routing applications.

Lawn Care Chat Query Router System Architecture



12 Key ML/NLP Features in this Classification System

- Transformer-Based Embeddings - Utilizes state-of-the-art transformer models (BERT, MiniLM) to create semantic vector representations of text.
- Zero-Shot Classification - Classifies queries into categories without explicit training on labeled examples, using semantic similarity instead.
- Hierarchical Classification - Implements a two-level classification approach with main categories and subcategories.
- Model Benchmarking - Compares performance across multiple embedding models (bert-base-uncased, bert-mini, MiniLM).
- Confusion Matrix Analysis - Visualizes classification errors to identify patterns in misclassifications.
- Performance Metrics Calculation - Calculates comprehensive metrics including accuracy, precision, recall, and F1 scores.
- Semantic Similarity Detection - Identifies semantically similar categories that might be redundant or confusing.
- Inference Time Measurement - Tracks query processing speed, enabling latency-based model selection.
- Error Analysis - Separately analyzes category and subcategory errors for targeted improvements.
- Alternative Prediction Tracking - Records second-best matches to evaluate classification confidence.
- Cosine Distance Calculation - Measures the "confidence gap" between best and second-best predictions.
- Topic Vector Similarity Analysis - Evaluates how close different category vectors are in the embedding space.

Python libraries and models used for this ML/NLP classification job:

Python Libraries:

- sentence-transformers - For generating embeddings and semantic similarity
- scikit-learn - For metrics, confusion matrices, and evaluation
- pandas - For data manipulation and management
- numpy - For numerical operations and array handling
- matplotlib/seaborn - For visualization and confusion matrices
- yaml - For configuration file parsing
- docx (python-docx) - For report generation
- openpyxl/xlsxwriter - For Excel report generation

ML/NLP Models:

- BERT-base-uncased - A general-purpose transformer model
- BERT-mini (prajjwal1/bert-mini) - A smaller, faster version of BERT

- MiniLM (all-MiniLM-L6-v2) - A distilled model optimized for sentence embeddings

The core ML/NLP approach uses a semantic similarity technique where:

- Topic descriptions are encoded into vector embeddings
- User queries are similarly encoded
- Cosine similarity is calculated between queries and topics
- The most similar topic is assigned as the classification

This approach is efficient for multi-class text classification without requiring large amounts of labeled training data, as it leverages pre-trained transformer models.

Topic Classification Structure

Loaded 38 topics from topics.yaml

Model: bert-base-uncased

Close Categories Analysis

Using a similarity threshold of 90% to identify closely related categories.

Found 11 pairs of closely related categories.

95.1% @@ Yard guidance/Vertebrate pests ## Yard guidance/Yard pests

Animals or pests that damage lawns, moles, voles, rabbits, deer, how to address vertebrate pests %% Pests like ants, mosquitoes, ticks, fleas, how to address yard pests that affect humans or pets

93.6% @@ Yard guidance/Insects ## Yard guidance/Yard pests

Insect pests of lawns or gardens, grubs, armyworms, chinch bugs, how to address insect issues %% Pests like ants, mosquitoes, ticks, fleas, how to address yard pests that affect humans or pets

93.5% @@ Yard guidance/Insects ## Yard guidance/Vertebrate pests

Insect pests of lawns or gardens, grubs, armyworms, chinch bugs, how to address insect issues %% Animals or pests that damage lawns, moles, voles, rabbits, deer, how to address vertebrate pests

92.6% @@ Yard guidance/Spots or patches ## Yard guidance/Disease

How to address spots or patches in the lawn, brown or yellow patches, thinning grass, dying grass, bare spots %% Lawn or garden diseases, large patch or brown patch, mushrooms in lawns, how to address or prevent diseases

92.3% @@ Yard guidance/Seeding ## Yard guidance/Weed control

When to seed lawn, best time to seed, unique seeding dates for the user, choosing the right grass seed for the lawn, recommendations for seed, best grass or seed types for lawn, identify grass in the lawn, seeding preparation, how to seed, step by step how to seed, seeding best practices, care after seeding to be successful, troubleshooting and problem solving seeding and germination issues, poor seed growth, conditions affecting seeding success %% When to treat weeds, best time to treat weeds, unique weed treatment dates for the user, timing weed control with weather, choosing the right weed control product, preemergent weed control, recommendations for weed control, best weed control products for lawn, how to eliminate or prevent specific weeds or weeds in general, how to address weed issues, identifying weeds in the lawn, what are the weeds in the lawn, how to identify weeds, note weeds includes broadleaves, sedges, grass weeds, and moss

91.8% @@ Yard guidance/Insects ## Yard guidance/Disease

Insect pests of lawns or gardens, grubs, armyworms, chinch bugs, how to address insect issues %% Lawn or garden diseases, large patch or brown patch, mushrooms in lawns, how to address or prevent diseases

90.8% @@ Yard guidance/Weed control ## Sunday products/Instructions

When to treat weeds, best time to treat weeds, unique weed treatment dates for the user, timing weed control with weather, choosing the right weed control product, preemergent weed control, recommendations for weed control, best weed control products for lawn, how to eliminate or prevent specific weeds or weeds in general, how to address weed issues, identifying weeds in the lawn, what are the weeds in the lawn, how to identify weeds, note weeds includes broadleaves, sedges, grass weeds, and moss %% What to do with received product, how to apply or use individual products, how to use sprayers with hose, how to use up the product in the container, how much a product covers, where to apply product, sequence of Sunday product applications, order or sequence or steps of using provided products like fertilizer (e.g. Grass Powerhouse, Fall Fortify, Potassium Boost, Core Seagreen), weed control (e.g. Dandelion Doom, Weed Warrior), or grass seed, how to use products in the right order, sequence of Sunday product applications with other lawn care practices, how to time Sunday products with other lawn care practices like watering, aerating, or mowing, using Sunday products with non-Sunday lawn and garden products

90.7% @@ Yard guidance/Weed control ## Yard guidance/Disease

When to treat weeds, best time to treat weeds, unique weed treatment dates for the user, timing weed control with

weather, choosing the right weed control product, preemergent weed control, recommendations for weed control, best weed control products for lawn, how to eliminate or prevent specific weeds or weeds in general, how to address weed issues, identifying weeds in the lawn, what are the weeds in the lawn, how to identify weeds, note weeds includes broadleaves, sedges, grass weeds, and moss %% Lawn or garden diseases, large patch or brown patch, mushrooms in lawns, how to address or prevent diseases

90.7% @@ Yard guidance/Weed control ## Yard guidance/Garden plant care

When to treat weeds, best time to treat weeds, unique weed treatment dates for the user, timing weed control with weather, choosing the right weed control product, preemergent weed control, recommendations for weed control, best weed control products for lawn, how to eliminate or prevent specific weeds or weeds in general, how to address weed issues, identifying weeds in the lawn, what are the weeds in the lawn, how to identify weeds, note weeds includes broadleaves, sedges, grass weeds, and moss %% Garden plants like vegetables, flowers, or shrubs, how to care for garden plants, how to grow vegetables or flowers, how to address issues with garden plants, how to grow a garden, how to care for a garden, how to grow specific plants in the garden

90.4% @@ Yard guidance/Disease ## Yard guidance/Grass types

Lawn or garden diseases, large patch or brown patch, mushrooms in lawns, how to address or prevent diseases %% Lawn grass identification, guidance for managing a specific lawn grass species, not grassy weeds

90.2% @@ Yard guidance/Soil problems ## Yard guidance/Disease

Soil quality or drainage issues, soil questions related to lawn health, hard or compacted soil %% Lawn or garden diseases, large patch or brown patch, mushrooms in lawns, how to address or prevent diseases

Detailed close category analysis saved to: ./Results/res_09/bert-base-uncased_close_categories.csv

Test Data

Loaded 485 test queries from Lawn Care Dataset - Clear Categories with Occasional Typos.csv

Classification Results

Time for first 10 queries: 0.8320 seconds

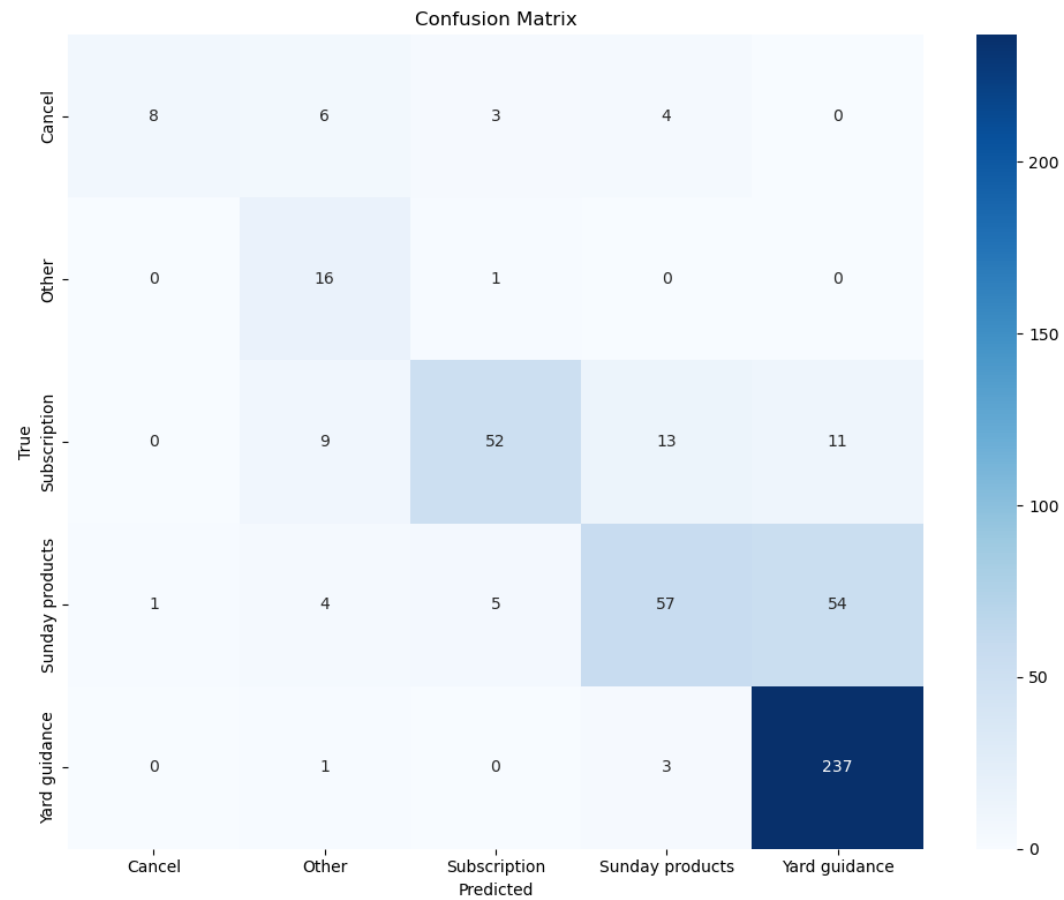
Time for last 10 queries: 0.8041 seconds

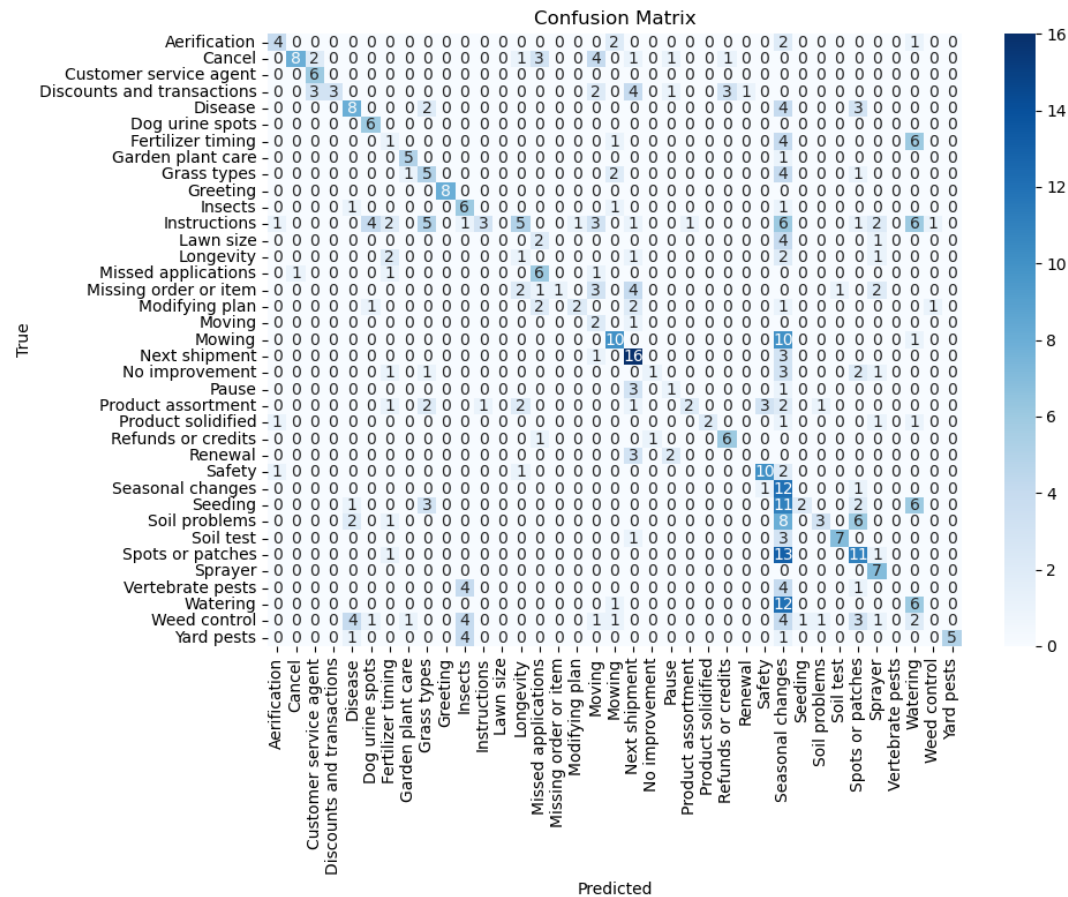
Average time per query: 0.0818 seconds

Queries per second: 12.22

Category accuracy: 76.29%

Category+Subcategory accuracy: 36.29%





Number of category errors: 115

Number of subcategory errors: 194

Number of correct classifications: 176

Detailed results exported to: ./Results/res_09/Results for model bert-base-uncased.xlsx

Model: prajjwal1/bert-mini

Close Categories Analysis

Using a similarity threshold of 90% to identify closely related categories.

Found 4 pairs of closely related categories.

94.9% @@ Yard guidance/Vertebrate pests ## Yard guidance/Yard pests

Animals or pests that damage lawns, moles, voles, rabbits, deer, how to address vertebrate pests %% Pests like ants, mosquitoes, ticks, fleas, how to address yard pests that affect humans or pets

94.1% @@ Yard guidance/Insects ## Yard guidance/Yard pests

Insect pests of lawns or gardens, grubs, armyworms, chinch bugs, how to address insect issues %% Pests like ants, mosquitoes, ticks, fleas, how to address yard pests that affect humans or pets

93.9% @@ Yard guidance/Insects ## Yard guidance/Vertebrate pests

Insect pests of lawns or gardens, grubs, armyworms, chinch bugs, how to address insect issues %% Animals or pests that damage lawns, moles, voles, rabbits, deer, how to address vertebrate pests

90.8% @@ Yard guidance/Weed control ## Sunday products/Instructions

When to treat weeds, best time to treat weeds, unique weed treatment dates for the user, timing weed control with weather, choosing the right weed control product, preemergent weed control, recommendations for weed control, best weed control products for lawn, how to eliminate or prevent specific weeds or weeds in general, how to address weed issues, identifying weeds in the lawn, what are the weeds in the lawn, how to identify weeds, note weeds includes broadleaves, sedges, grass weeds, and moss %% What to do with received product, how to apply or use individual products, how to use sprayers with hose, how to use up the product in the container, how much a product covers, where to apply product, sequence of Sunday product applications, order or sequence or steps of using provided products like fertilizer (e.g. Grass Powerhouse, Fall Fortify, Potassium Boost, Core Seagreen), weed control (e.g. Dandelion Doom, Weed Warrior), or grass seed, how to use products in the right order, sequence of Sunday product applications with other lawn care practices, how to time Sunday products with other lawn care practices like watering, aerating, or mowing, using Sunday products with non-Sunday lawn and garden products

Detailed close category analysis saved to: ./Results/res_09/prajjwal1-bert-mini_close_categories.csv

Test Data

Loaded 485 test queries from Lawn Care Dataset - Clear Categories with Occasional Typos.csv

Classification Results

Time for first 10 queries: 0.2396 seconds

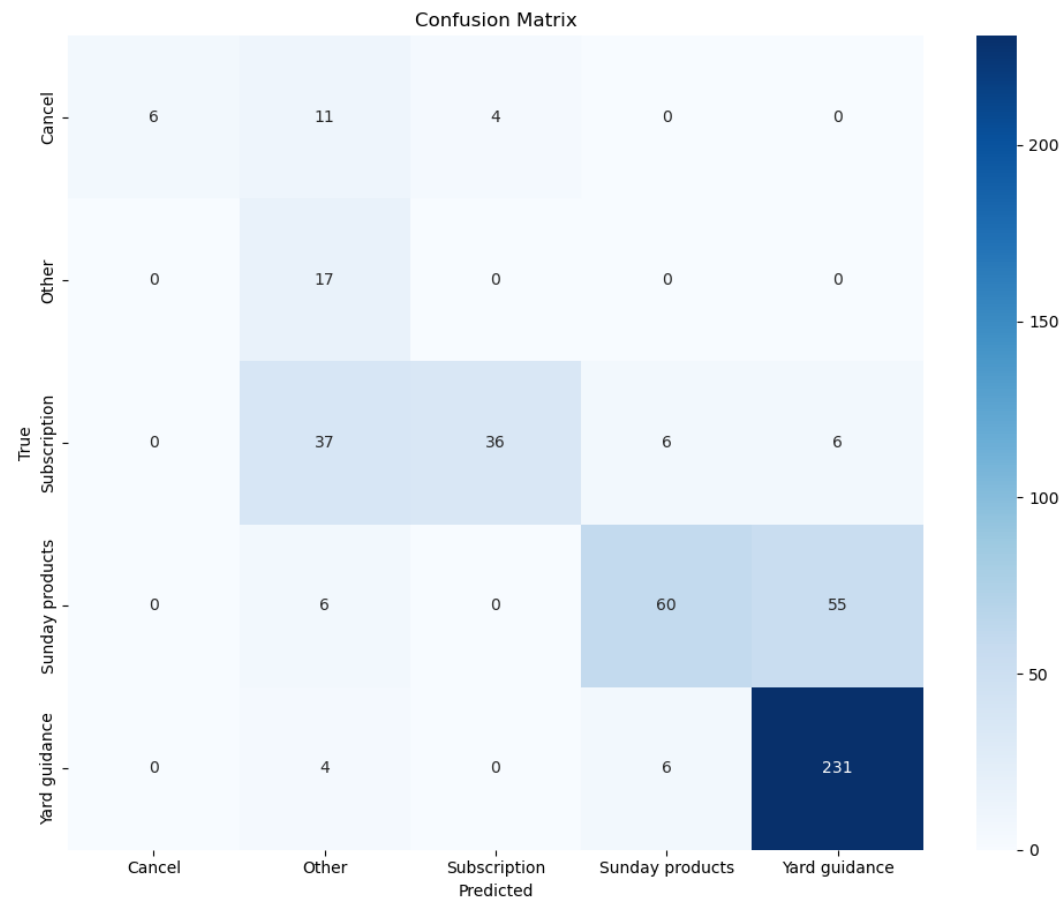
Time for last 10 queries: 0.2163 seconds

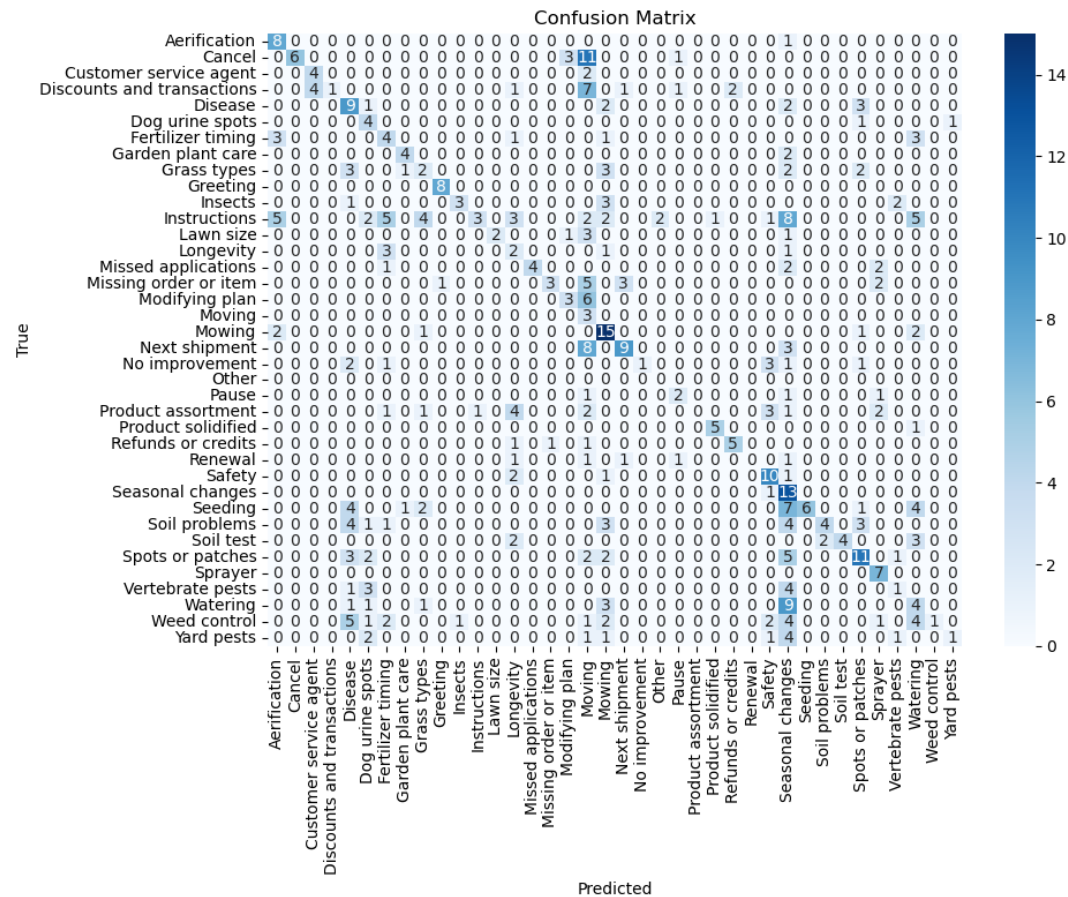
Average time per query: 0.0228 seconds

Queries per second: 43.87

Category accuracy: 72.16%

Category+Subcategory accuracy: 35.46%





Model: all-MiniLM-L6-v2

Close Categories Analysis

Using a similarity threshold of 90% to identify closely related categories.

No closely related categories found with the current threshold.

Consider lowering the threshold if you expect to find similar categories.

Test Data

Loaded 485 test queries from Lawn Care Dataset - Clear Categories with Occasional Typos.csv

Classification Results

Time for first 10 queries: 0.2367 seconds

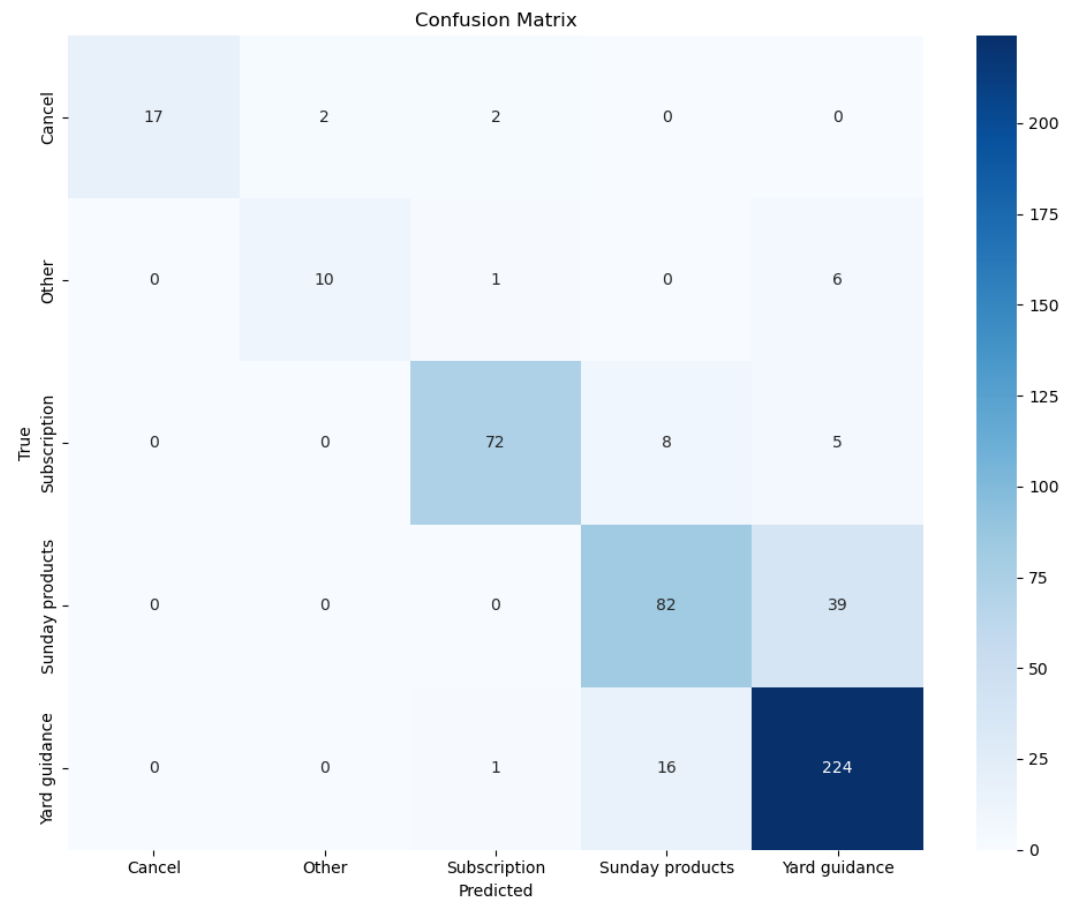
Time for last 10 queries: 0.2336 seconds

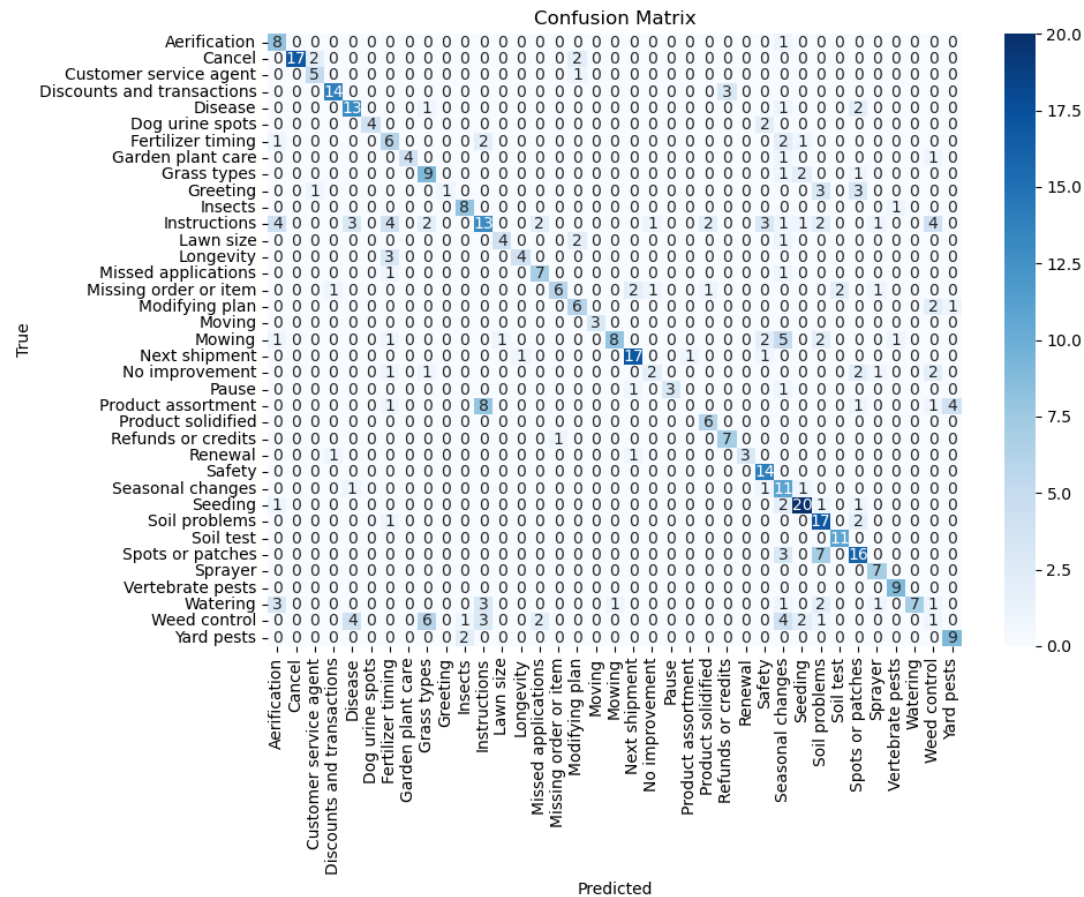
Average time per query: 0.0235 seconds

Queries per second: 42.52

Category accuracy: 83.51%

Category+Subcategory accuracy: 61.86%





Number of category errors: 80

Number of subcategory errors: 105

Number of correct classifications: 300

Detailed results exported to: ./Results/res_09/Results for model all-MiniLM-L6-v2.xlsx

Category-Level Model Comparison

This comparison shows model performance at the CATEGORY level only (ignoring subcategories):

Model	Accuracy	Precision	Recall	F1 Score	Queries/sec	Time/query
bert-base-uncased	76.29%	77.81%	76.29%	74.65%	12.22	81.81 ms
prajjwal1/bert-mini	72.16%	81.00%	72.16%	71.90%	43.87	22.80 ms
all-MiniLM-L6-v2	83.51%	83.78%	83.51%	83.22%	42.52	23.52 ms

Category+Subcategory-Level Model Comparison

This comparison shows model performance at the CATEGORY+SUBCATEGORY level (both must be correct):

Model	Accuracy	Precision	Recall	F1 Score	Queries/sec	Time/query
bert-base-uncased	36.29%	52.56%	36.29%	34.19%	12.22	81.81 ms
prajjwal1/bert-mini	35.46%	59.48%	35.46%	34.27%	43.87	22.80 ms
all-MiniLM-L6-v2	61.86%	64.13%	61.86%	59.52%	42.52	23.52 ms

Execution Summary

Total execution time: 105.01 seconds

Appendix

Classification Glossary

Accuracy - The overall proportion of correct predictions (both true positives and true negatives) over the total number of predictions.

Precision (Positive Predictive Value) - The proportion of true positive predictions among all positive predictions. Shows how reliable positive classifications are.

Recall (Sensitivity, True Positive Rate) - The proportion of actual positives that were correctly identified. Shows how well the classifier finds all positive cases.

F1 Score - The harmonic mean of precision and recall, providing a balance between the two, especially useful for imbalanced datasets.

Area Under ROC Curve (AUC-ROC) - Measures the classifier's ability to discriminate between classes across various thresholds.

Confusion Matrix - A table showing true positives, false positives, true negatives, and false negatives.

Specificity (True Negative Rate) - The proportion of actual negatives correctly identified.

False Positive Rate - The proportion of actual negatives incorrectly classified as positive.

Classification Threshold - The decision boundary used to determine class assignment.

Class Distribution - Information about the balance or imbalance of classes in your dataset.

Model Confidence/Probability Scores - The probability estimates for predictions rather than just the final classifications.

Training/Testing Performance Gap - Indicates potential overfitting or underfitting.

Feature Importance - Which features have the greatest impact on classification outcomes.

Cross-Validation Results - Performance metrics across multiple data splits.

Runtime/Inference Speed - How quickly the classifier makes predictions.