GROUP D

THE DOC-ASSISTANT

GROUP D

JHARANA ADHIKARI

UMA MAHESHWARI

RISHABH JHA

SHISHIR MISHRA

AGENDA

- INTRODUCTION
- MOTIVATION
- SYSTEM FUNCTIONALITY
- DATA SOURCE
- DATA PREPROCESSING
- EXPLORATORY DATA ANALYSIS
- TOPIC MODELING WITH LDA
- LDA TOPIC EXAMPLE
- LLM USING BERT
- RESULT
- CONCLUSION
- FUTURE WORK
- REFERENCE

INTRODUCTION TO AI IN HEALTHCARE

- The advent of Artificial Intelligence (AI) has significantly transformed various business sectors.
- The healthcare and medical sciences domain has notably benefited from AI advancements.
- AI has accelerated biomedical research, leading to rapid developments such as the creation of COVID-19 vaccines and breakthroughs in cancer and neuroscience research.
- Despite these advancements, day-to-day healthcare operations still have considerable scope for improvement through AI integration.

MOTIVATION

- Daily operations at healthcare centers are often inefficient and costly.
- Patients spend unnecessary time and money to determine which specialty service they need.
- Healthcare providers face challenges in managing time and resources to cover more patients effectively.
- AI can streamline these processes, improving patient care and resource management.

SYSTEM FUNCTIONALITY

The Medical Assistant system has two primary functions:

- **Symptom Analysis:** The system analyzes patient symptoms to suggest appropriate medical services, saving time and money for both patients and providers.

- **Drug Recommendation:** The system assists healthcare providers by recommending drugs, aiding doctors and paramedics in covering more patients efficiently.

Due to FDA and WHO guidelines, drug recommendations are directed to healthcare providers rather than patients directly.

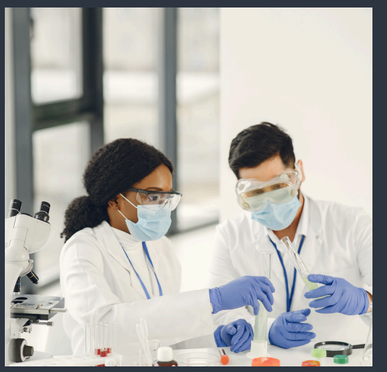
DATA SOURCE

The data is sourced from Drugs.com, a comprehensive database of prescription drugs.

- The dataset is stored in a JSONL (JSON Lines) file format, where each line is a separate JSON object.
- Due to the large size of the data, it is loaded in chunks to prevent memory issues.

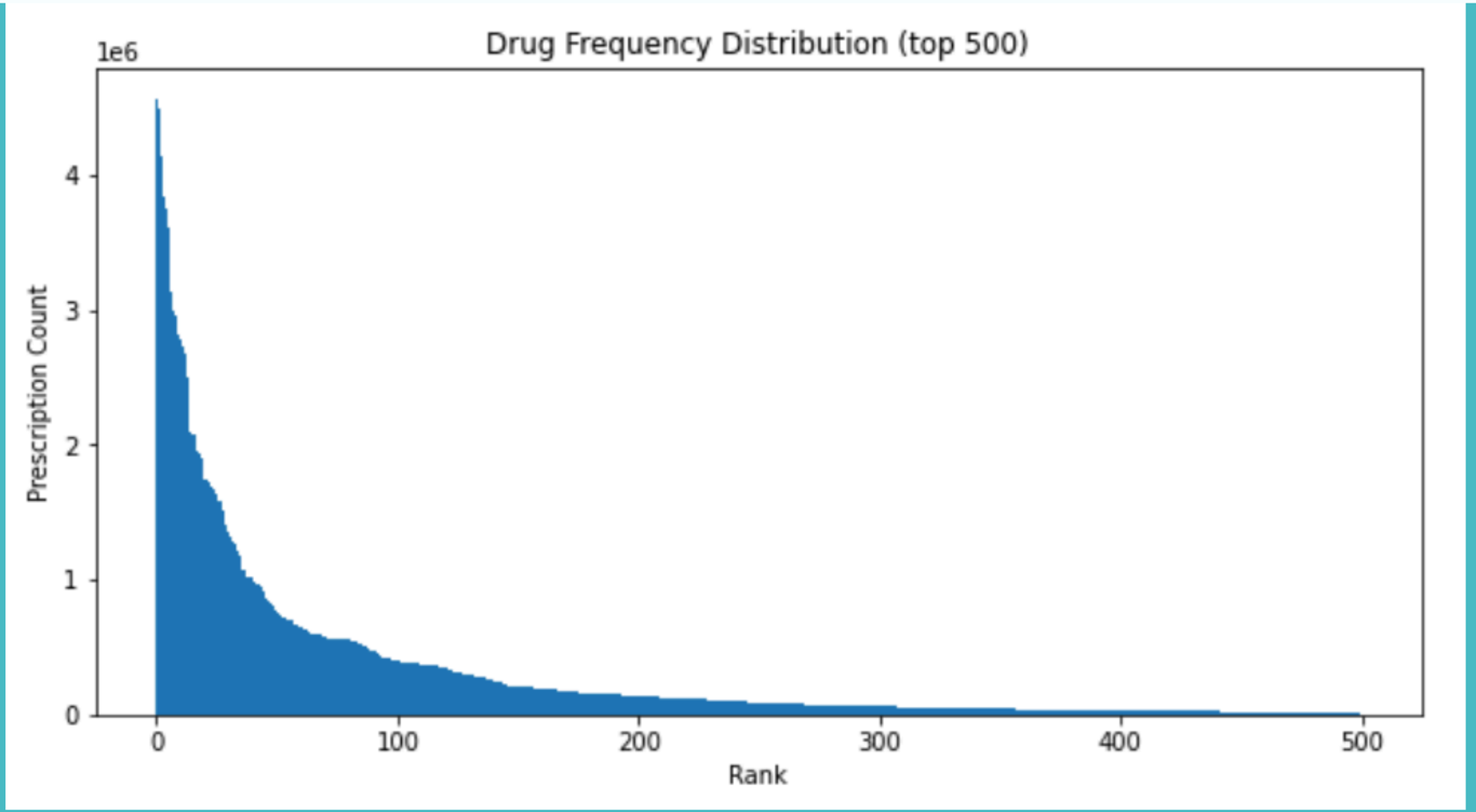
DATA PREPROCESSING

- Data is sourced from Drugs.com in JSONL format.
- Use pd.read_json with lines=True for efficient loading.
- Filter out providers with fewer than 40 unique drugs and specialties with fewer than 40 providers.
- Expand provider_variables into a separate data frame using json_normalize.



EXPLORATORY ANALYSIS

- The dataset contains 29 unique classes related to different medical specialties.
- There is a significant class imbalance, with some classes representing a large portion of the data while others are underrepresented.
- Understanding this distribution is crucial for building effective models.



Class distribution in Bar chart

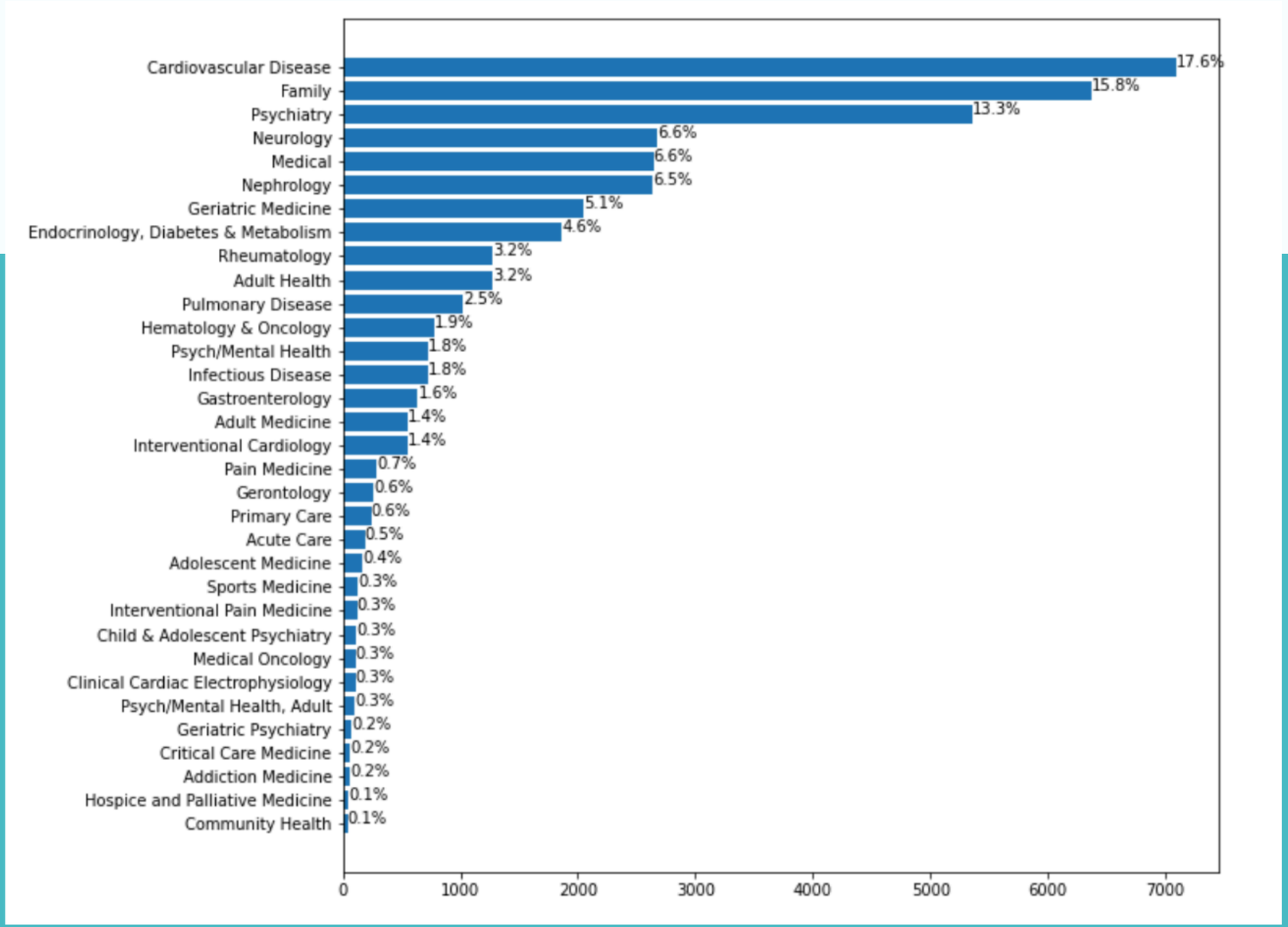
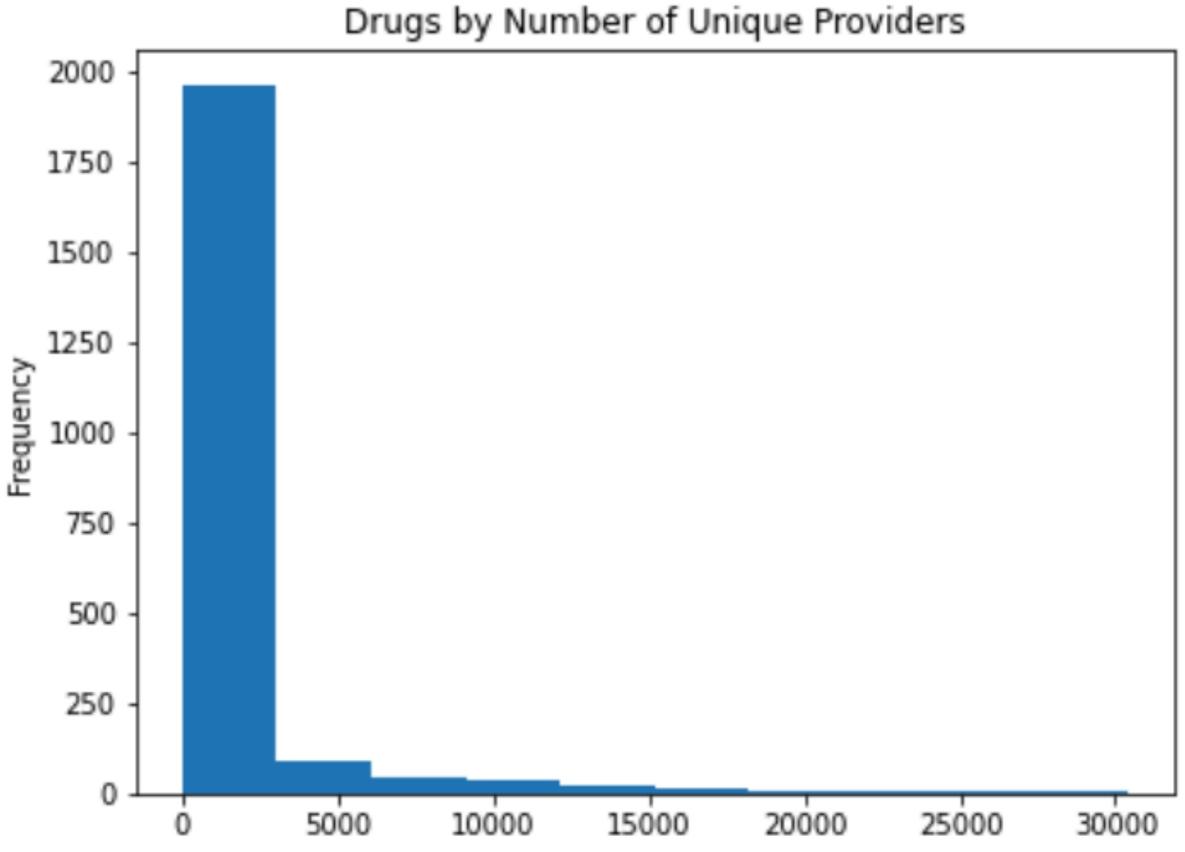
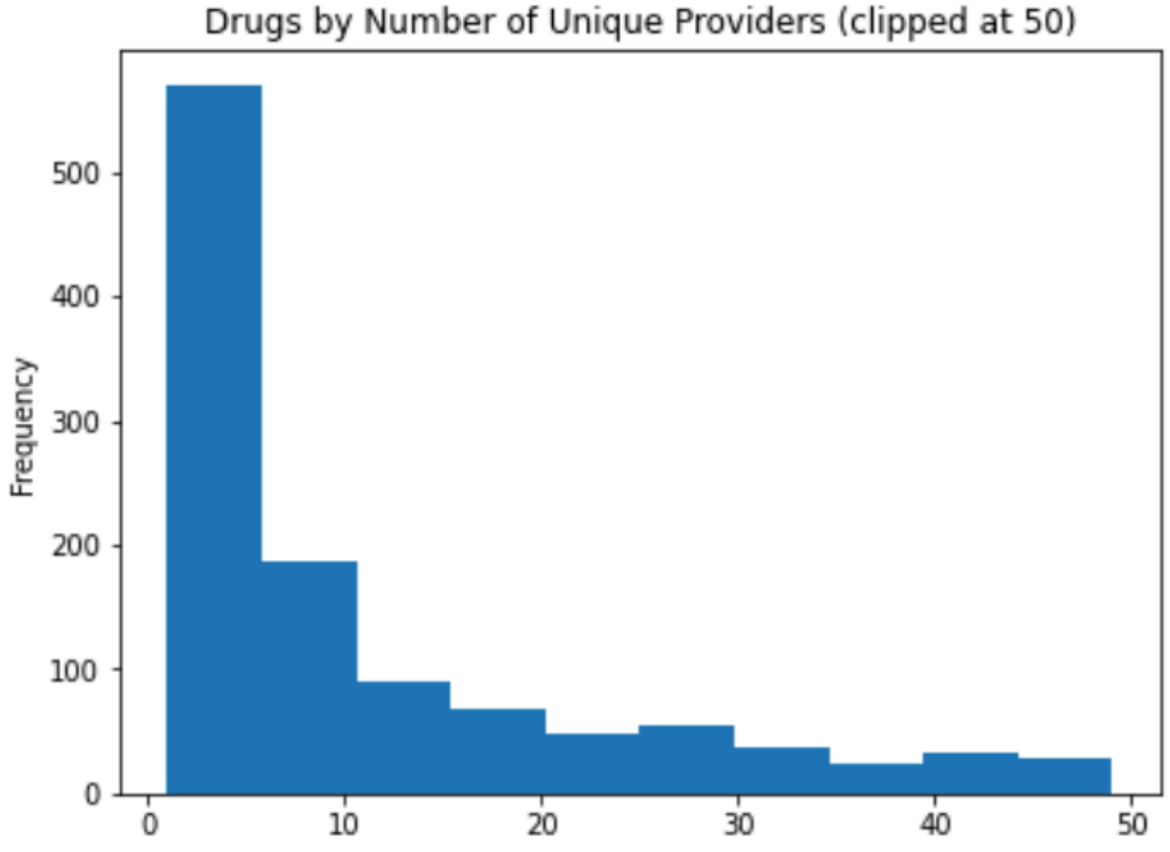


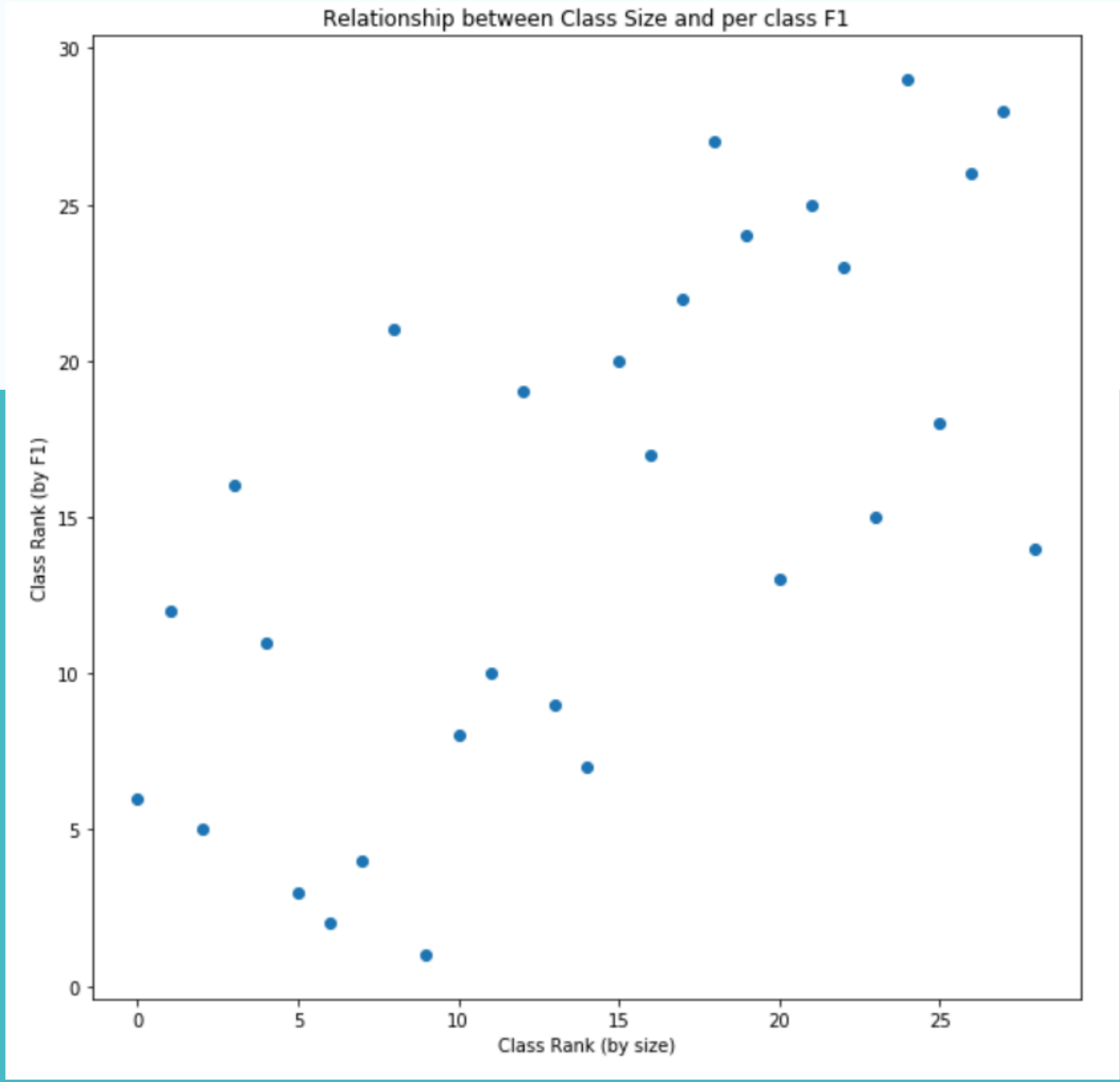
Table of Top 15 Drug Brands with Correlation Insights

	Drug Name	Count	Proportion of All Prescriptions	Cumulative Proportion of All Prescriptions
0	LISINOPRIL	4563509	0.027283	0.027283
1	AMLODIPINE BESYLATE	4494231	0.026869	0.054152
2	SIMVASTATIN	4144629	0.024779	0.078931
3	FUROSEMIDE	3834944	0.022927	0.101858
4	ATORVASTATIN CALCIUM	3750135	0.022420	0.124278
5	LEVOTHYROXINE SODIUM	3613821	0.021605	0.145883
6	METOPROLOL TARTRATE	3136307	0.018750	0.164634
7	HYDROCODONE-ACETAMINOPHEN	2985613	0.017850	0.182483
8	OMEPRAZOLE	2966475	0.017735	0.200218
9	CLOPIDOGREL	2809993	0.016800	0.217018
10	METOPROLOL SUCCINATE	2776497	0.016599	0.233617
11	GABAPENTIN	2726888	0.016303	0.249920
12	CARVEDILOL	2671914	0.015974	0.265894
13	WARFARIN SODIUM	2498211	0.014936	0.280830
14	CLONAZEPAM	2090653	0.012499	0.293329

Drug Brands and Unique Providers Overview

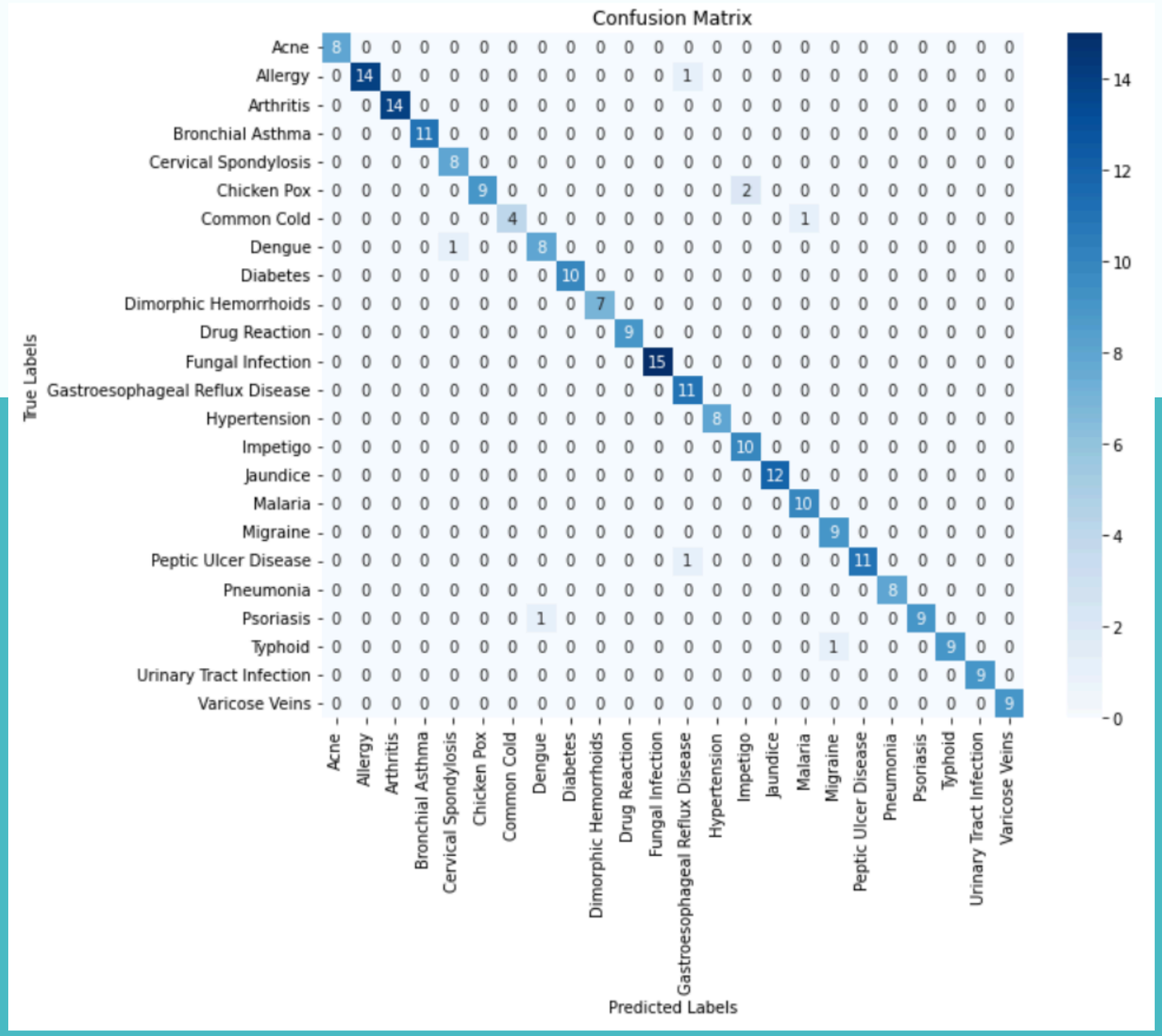






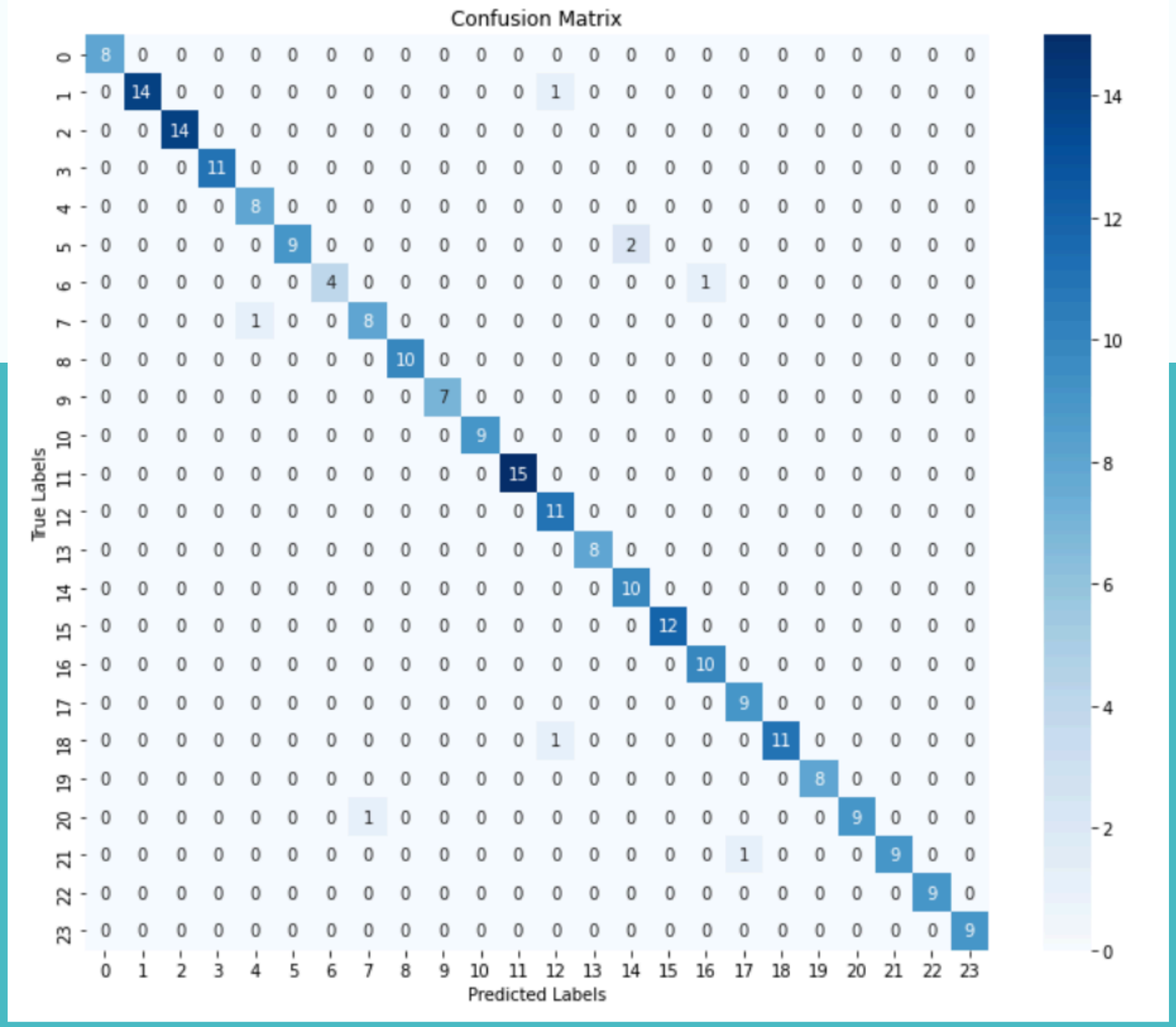
Class Size vs. F1 Score: Analyzing Performance Patterns

- This scatter plot shows how class size relates to F1 scores.
- It reveals that low performance is often due to confusion between similar classes, such as "Child & Adolescent Psychiatry" vs. "Psychiatry" and "Acute Care" vs. "Family," rather than small sample sizes.



Confusion Matrix Heatmap Visualization for Model Prediction

- The confusion matrix illustrates the model's performance in disease classification, showing high accuracy across many categories.
- Most diseases are predicted perfectly, with only minor misclassifications in a few cases, indicating a robust model with high precision and recall overall.



Classification Metrics for Disease Classification Model

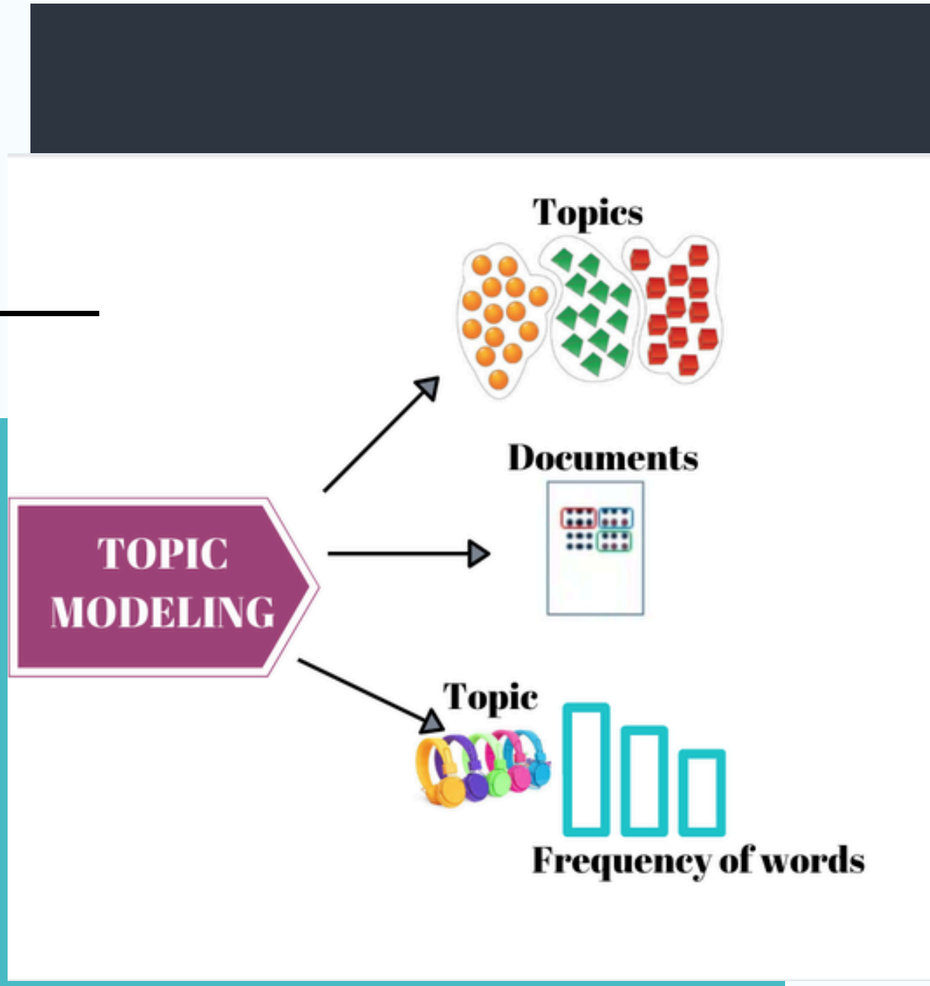
- The model demonstrates high overall correctness with an accuracy of 0.9667 and a weighted precision of 0.9706.
- It shows high precision and recall for most classes, with only minor deviations in a few categories, indicating its reliability for practical disease diagnosis.

TOPIC MODELING WITH LDA

- **Latent Dirichlet Allocation (LDA):** An unsupervised technique to identify topics (groups of related words) in text data.

- **Application:** Each provider can be seen as a document and each drug prescribed as a word in that document.

- **Purpose:** LDA helps in dimensionality reduction and denoising features by identifying related drug groups.



LDA with Logistic Regression Topic Examples

Topic 4

TEMAZEPAM, LORAZEPAM, CLONAZEPAM (Sedatives)

Topic 7

COPAXONE, BACLOFEN, AVONEX (MS medications)

Topic 10

NORVIR, TRUVADA, ISENTRESS (HIV medications)

Topic 13

MORPHINE SULFATE ER, OXYCODONE HCL, OXYCODONE HCL-ACETAMINOPHEN (Opiate pain relievers)

Topic 28

JANUMET, TRADJENTA, JANUVIA (Diabetes medications)

Topic 47

LEVETIRACETAM, VIMPAT, LAMOTRIGINE (Seizure medications)

LARGE LANGUAGE MODELLING USING BERT

- **Preprocessing:** Used BERT embeddings for patient symptoms and drug descriptions to capture semantic relationships.
- **Fine-Tuning**: Adapted BERT for symptom-to-drug classification to enhance prediction accuracy.
- **Imbalanced Data:** Applied techniques to handle imbalanced classes and improve model performance.
- **Performance Evaluation:** Compared BERT’s F1 scores to traditional methods, highlighting its effectiveness in drug recommendation.

CONCLUSION

Summary:

The project successfully developed a Medical Assistant system that analyzes symptoms and recommends drugs.

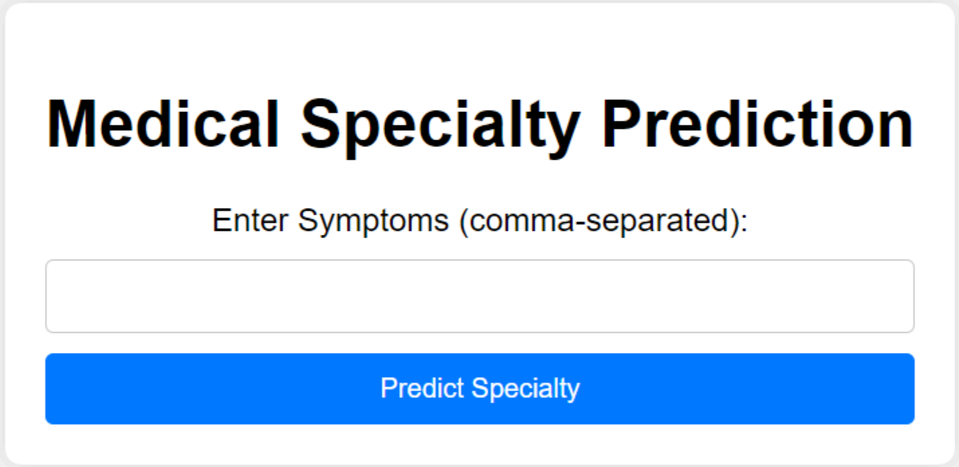
Findings:

Imbalanced class distribution posed challenges; topic modeling and advanced classification techniques were utilized.

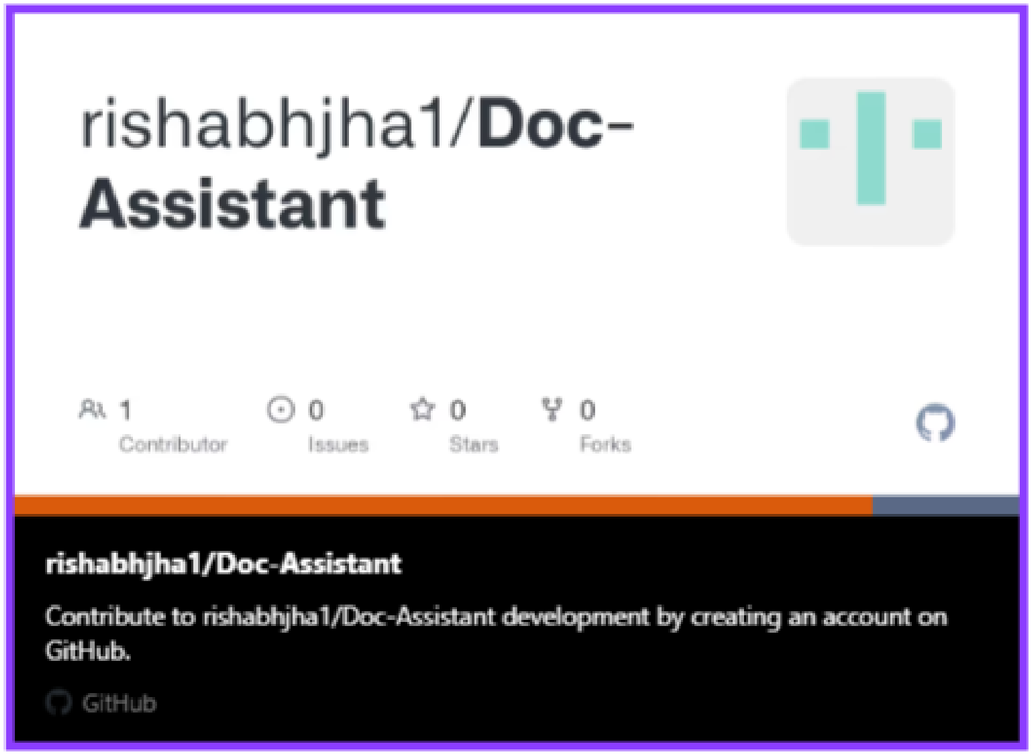
FUTURE WORK

- Make our own LLM to implement the system
- Enhancing model performance for low-frequency classes.
- Further refining drug recommendation algorithms to assist healthcare providers more effectively.

Front End with Flask



https://github.com/rishabhjha1/Doc-Assistant



References

- Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J. 2021 Jul;8(2):e188-e194. doi: 10.7861/fhj.2021-0095. PMID: 34286183; PMCID: PMC8285156.

- Saskia Locke, Anthony Bashall, Sarah Al-Adely, John Moore, Anthony Wilson, Gareth B. Kitchen, Natural language processing in medicine: A review, Trends in Anaesthesia and Critical Care,Volume 38, 2021,Pages 4-9, ISSN 2210-8440, https://doi.org/10.1016/j.tacc.2021.02.007.

- Hao T, Huang Z, Liang L, Weng H, Tang B. Health Natural Language Processing: Methodology Development and Applications. JMIR Med Inform. 2021 Oct 21;9(10):e23898. doi: 10.2196/23898. PMID: 34673533; PMCID: PMC8569540.

- Tamang S, Humbert-Droz M, Gianfrancesco M, Izadi Z, Schmajuk G, Yazdany J. Practical Considerations for Developing Clinical Natural Language Processing Systems for Population Health Management and Measurement. JMIR Med Inform. 2023 Jan 3;11:e37805. doi: 10.2196/37805. PMID: 36595345; PMCID: PMC9846439.

- Yadav, S. K., & Singh, A. K. (2022). Applications of artificial intelligence in medical imaging: Current trends and future directions. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-022-13428-4

Thank You