**Date:12/04/2025**

**DATA SCIENCE & DATA ANALYTICS|AI|ML**

**PROJECT 2 REPORT**

A local retail pharmacy called Pharma Lu has a database where they put data on stock received and dispensed to patients. Lately the pharmacy has been having issues with students mistakenly giving the patients prescription medicines instead of over the counter medicine. The pharmacy would like to have a dispensing system that can highlight to anyone dispensing medication if they are dispensing a prescription medicine to avoid any errors being made. Students will use the dispensing system that will be connected to the database to dispense medication. The dispensing system should be able to identify whether it's a prescription or an over the counter medication.

A machine learning model will be created that will consider the features of a medicine (e.g., Category, Dosage Form, Manufacturer), and predict whether it is a prescription or over the counter medication. This model will form a part of the dispensing system.

The following are the research questions that were answered through the exploratory data analysis:

Q1. Which dosage form is more prominent in the prescription vs over the counter category?

Q2:What are the top most used medications in prescription and over the counter?

Q3: What are the most frequent dosage forms in prescription vs over the counter?

Q4: What are the top 5 manufacturers for prescription medicine?

Q5. Which category of medicines commonly falls under prescription medicine?

Q6. What is the proportion of medicines prescription vs over the counter?

Q7. Is there evidence of market dominance by particular manufacturers?

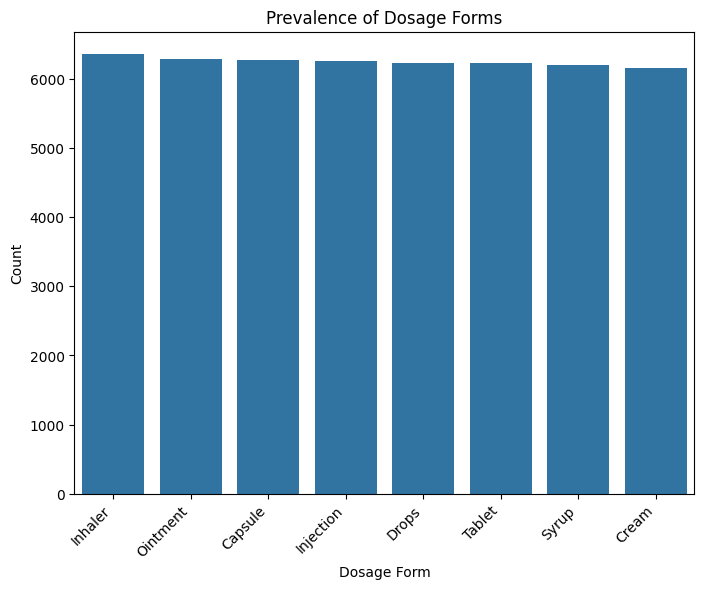
Q8. How does the distribution of manufacturers differ between prescription and over-the-counter medications?

Q9. What are the top 10 most frequent medication categories and classification of prescription or over the counter?

Q10.How does the frequency of individual medications differ between prescription and over-the-counter classifications?

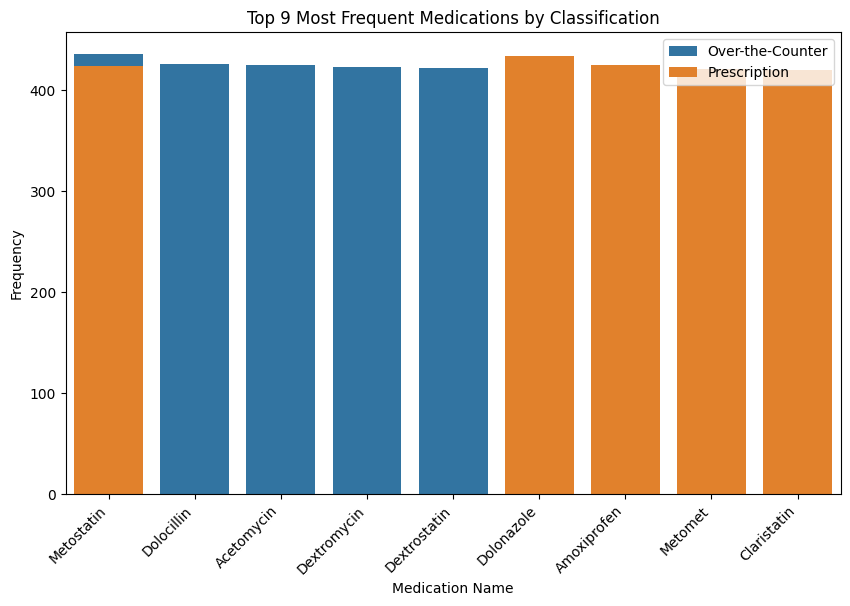
**Exploratory Data Analysis (EDA)**

The EDA phase involved a comprehensive analysis of the provided loan data. Key visualizations and their insights are summarized below:



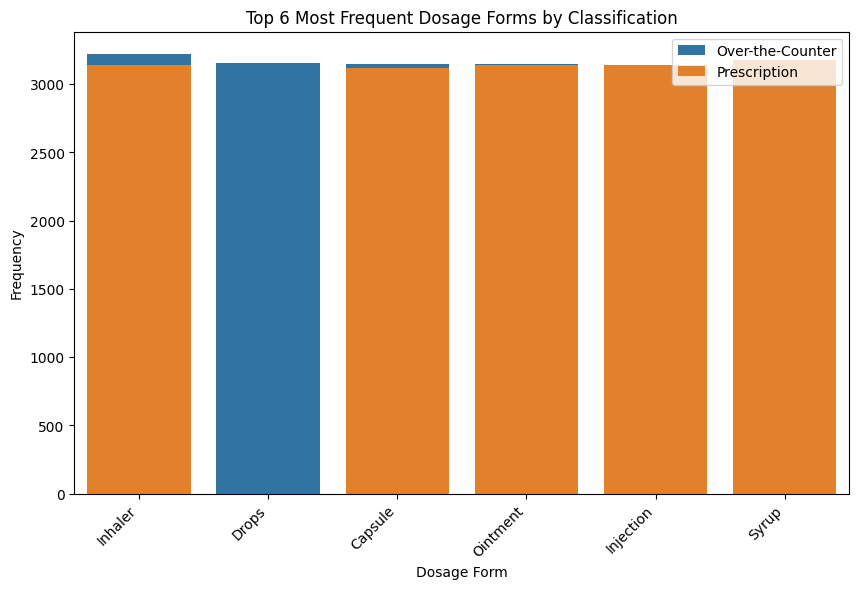
**Figure 1: Dosage Form Prevalence**

There is a more significant difference in count with the syrup. More of the syrups were prescription as compared to over the counter. Inhalers and drops are mostly over the counter as compared to prescription.



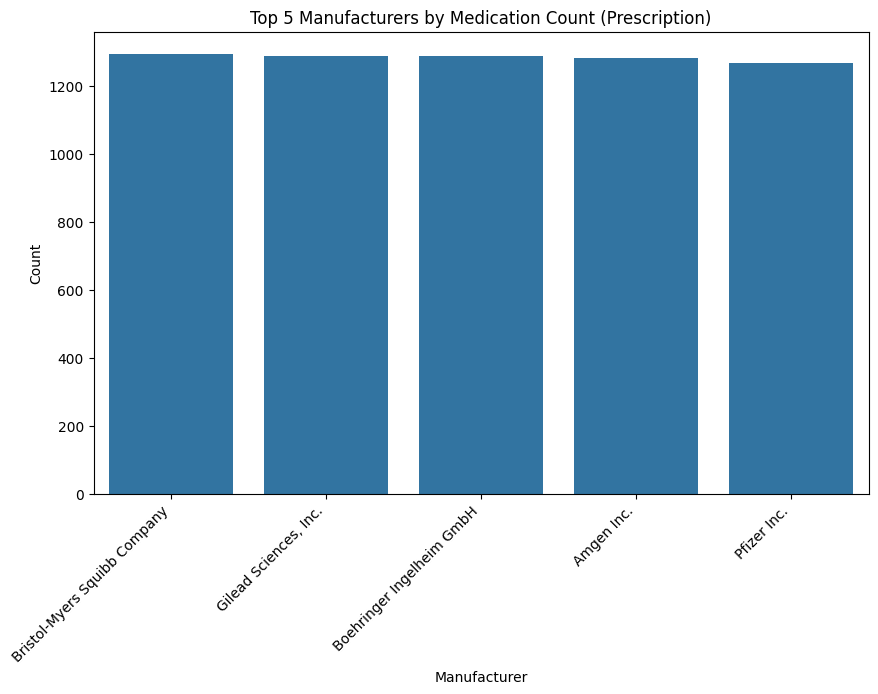
**Figure 2: Top 9 Most Frequent Medications by Classification**

Dolocillin, Acetomycin, Dextromycin and desxtrostatin are the most used over the counter medicines whilst Dolonazole is the most used prescription medication.



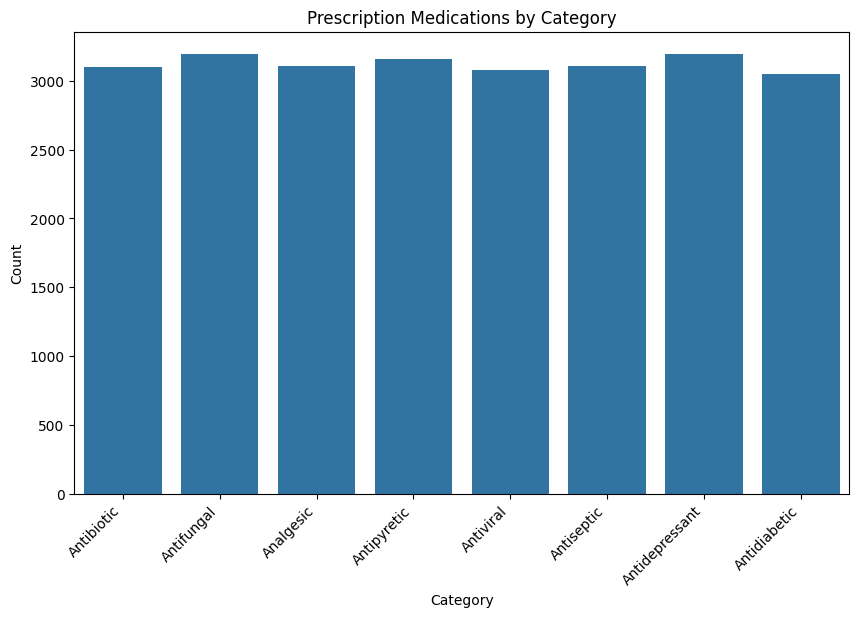
**Figure 3: Top 6 frequent dosage forms classification**

Drops are the most used dosage form for over the counter medicines, followed by inhaler and capsule. Syrup, injection, and ointment are the most used for prescription medicine.



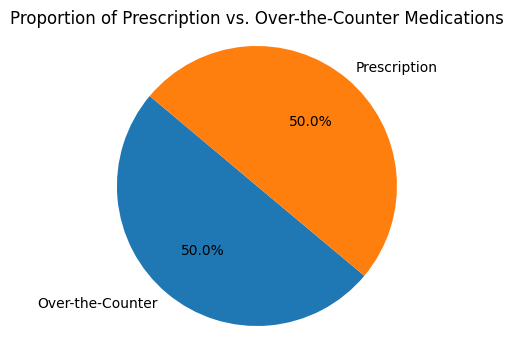
**Figure 4: Top 5 manufacturers for prescription medicine**

The top manufacturer for prescription medicine is Bristol-Myers Squibb Company which has a slightly higher count of medicines supplied from them as compared to the other manufacturers.



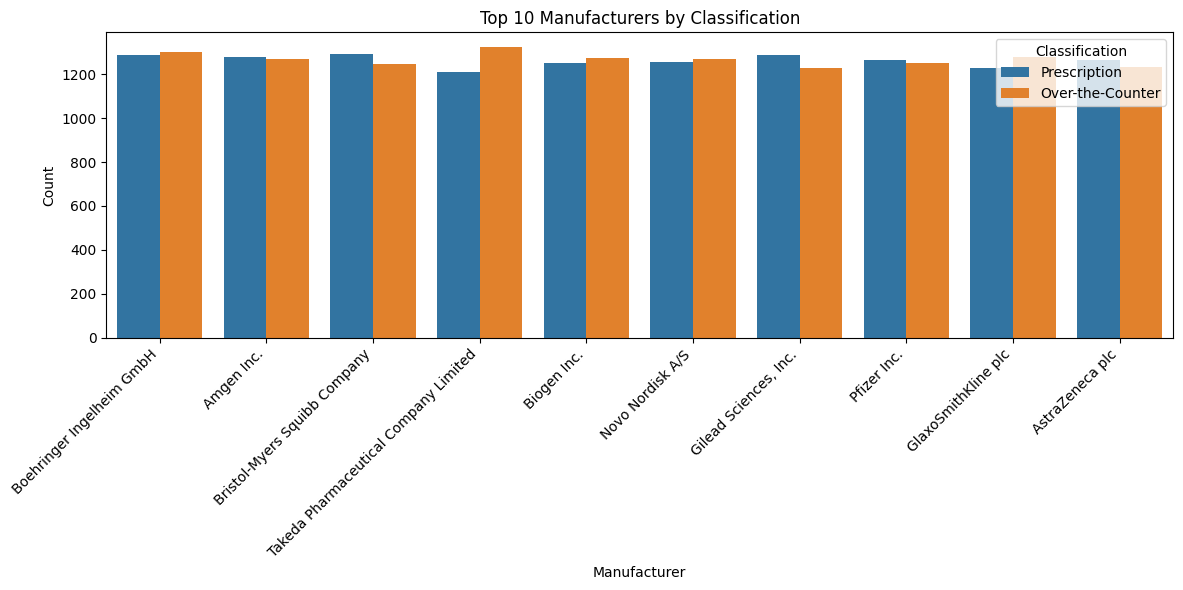
**Figure 5:Prescription Medications by Category**

Medicines that fall in the prescription category are antibiotic, antifungal, analgesic, antipyretic, antiviral, antiseptic,antiviral, antiseptic antidepressants and antidiabetic. Antifungal. antiviral and antidepressant have the highest count.



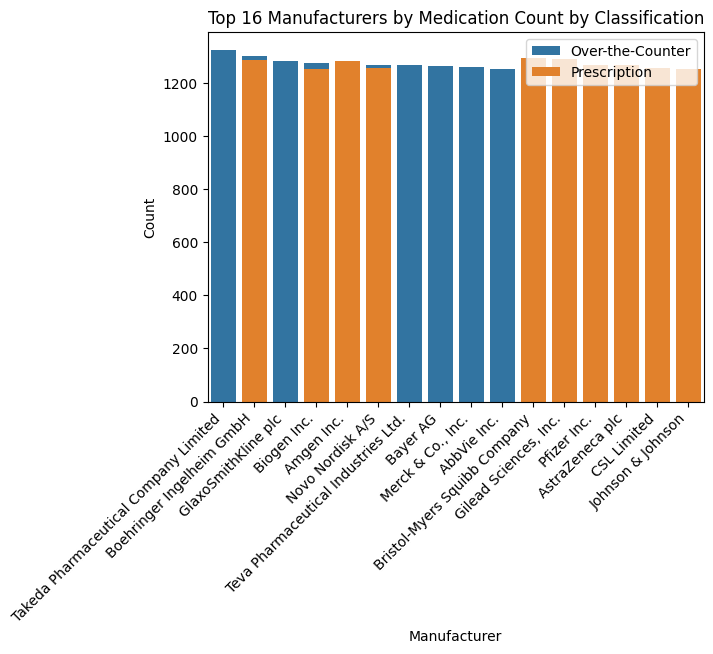
**Figure 6:Proportion of prescription medications vs over the counter medicines**

There is an equal amount of prescription and over the counter medicines that are stocked. There is a 50% chance of the student dispensing either a prescription medicine or an over the counter medicine.

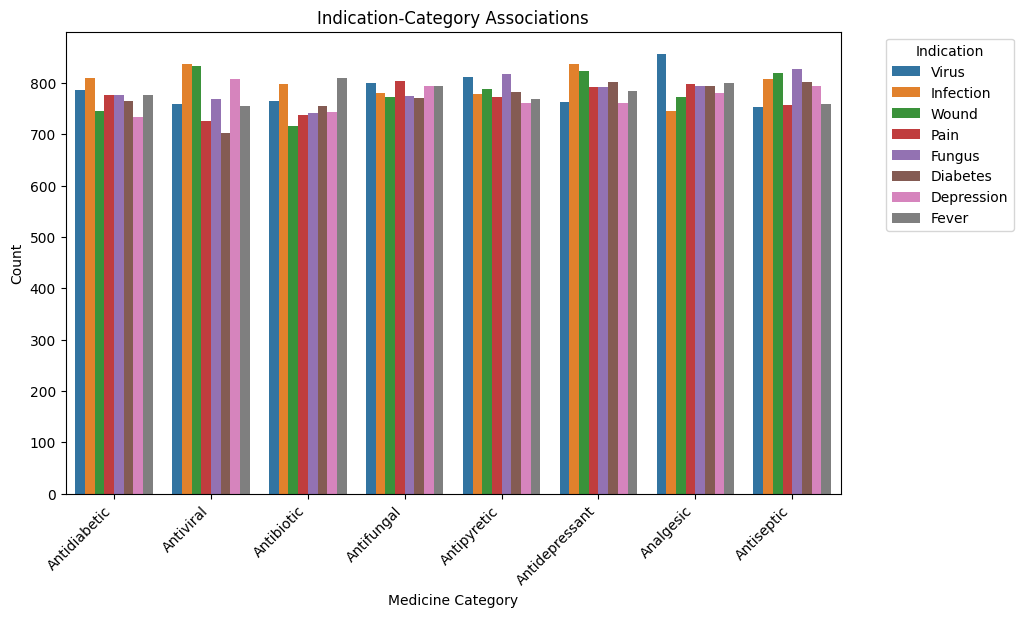
****

**Figure 7:Market dominance by particular manufacturers**

All manufacturers have supplied an amount of medicine within a similar range all above a count of 2500. Boehringeer Ingelheim GmbH had a slightly higher count of medicine produced compared to everyone else.All the manufacturers have similar amounts of prescription vs over the counter. Takeda pharmaceutical has a slightly higher over the counter which means if a medicine from Takeda pharmaceutical is being dispensed then it may most likely be over the counter.

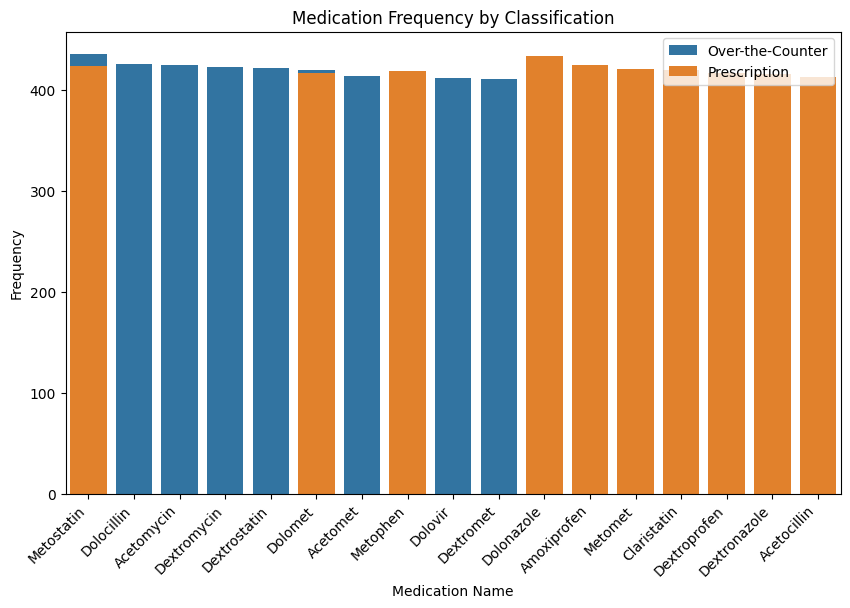


**Figure 8:Manufacturer Distribution by Classification**  
Amongst the top 10 62.5% of the manufacturers manufacture medicines that are prescription medicines. The count of medicines from all manufacturers are above 1200.Takeda pharmaceutical company limited has the highest medicine count for over the counter medication. There is not much difference between the amount of medicine from all manufacturers.



**Figure 9:Relationship between indication and categor**y

Each category of medicine can be given to treat different indications which are infection, wound, pain, fungus, diabetes,depression or fever. The prescription medicines that are most frequently dispensed are analgesic, antiseptic,antidepressant, antidiabetic, antiviral, antifungal and antibiotic. Analgesics given for have the highest count which means they are dispensed the most especially for viruses.Antivirals have the second highest count for being given for infections and wounds.



**Figure 10: Medication frequency by classification**

10 of the medicines in the graph are prescription medicines except for Metostatin and Dolomet which can be either prescription or Metostatin and Dolomet are most frequently given as an over the counter medicine.However Whereas only 7 medicines are over the counter. All medicines have a frequency above 400 with Dolanozole and Metostatin have the highest frequency.

Findings from the descriptive statistics show that there are 64 unique medication names, suggesting some medications are very frequently repeated."Metostatin" is the most frequent medication name, appearing 860 times. This indicates it is a commonly dispensed medication."Antidepressant" is the most frequent category, appearing 6354 times. There is also a very large number of unique strength values (999), which means that the strength variable is very granular."Boehringer Ingelheim GmbH" is the most frequent manufacturer, with 2587 medications. This suggests that this manufacturer is a significant supplier.Infection" is the most frequent indication, appearing 6393 times. This indicates that a large portion of the medicines are for infections.

**Machine learning model**

This model provides a probability of a medication being prescription or over-the-counter. The model utilizes a random forest classifier and leverages key features such as medication name, category, dosage form, and manufacturer. Label encoder was used to encode the categorical features 'Name', 'Category', 'Dosage Form', 'Manufacturer' into numerical values.The dataset was then split into training and testing sets (80/20).

The model's performance was evaluated using the following metrics:

* Accuracy: 0.49
* Precision: 0.49 (weighted avg)
* Recall: 0.49 (weighted avg)
* F1-score: 0.49(weighted avg)
* AUC-ROC: 0.49

**Inferential statistics**

For inferential statistics the chi square test was used as this test is suitable for categorical variables. This test was used to test whether the variables affect either the prediction of a medicine being either prescription or over the counter.

**These are the hypotheses for Individual Predictor Variables:**

**Name:**

**Null Hypothesis (H0):** The medication's name has no statistically significant effect on the probability of it being a prescription medication.

**Alternative Hypothesis (H1):** The medication's name has a statistically significant effect on the probability of it being a prescription medication.

**Category:**

**Null Hypothesis (H0):** The medication's category has no statistically significant effect on the probability of it being a prescription medication.

**Alternative Hypothesis (H1):** The medication's category has a statistically significant effect on the probability of it being a prescription medication.

**Dosage Form:**

**Null Hypothesis (H0)**: The medication's dosage form has no statistically significant effect on the probability of it being a prescription medication.

**Alternative Hypothesis (H1):** The medication's dosage form has a statistically significant effect on the probability of it being a prescription medication.

**Manufacturer:**

**Null Hypothesis (H0):** The medication's manufacturer has no statistically significant effect on the probability of it being a prescription medication.

**Alternative Hypothesis (H1):** The medication's manufacturer has a statistically significant effect on the probability of it being a prescription medication.

**Results of inferential statistics**

This means there is no statistically significant association between the medication's category and its classification (prescription vs. OTC).The category of the medication does not seem to influence if the medication is prescription or over the counter.There is no statistically significant association between the medication's dosage form and its classification.The dosage form of the medication does not seem to influence if the medication is prescription or over the counter.There is no statistically significant association between the medication's manufacturer and its classification.The manufacturer of the medication does not seem to influence if the medication is prescription or over the counter.The distribution of prescription vs over the counter medications is fairly even across all the tested features.This is very important information for the machine learning model, as it indicates that these features on their own are not strong predictors. Other features, or combinations of features, may be needed to build a strong predictive model.

**Summary**

Regarding dosage forms, syrups, injections, and ointments were frequent in prescriptions, while drops were the most common dosage form for over the counter medications. Bristol-Myers Squibb Company was identified as a leading manufacturer for prescription medications, though no significant market dominance by any single manufacturer was observed.Prescription medications commonly fell within categories such as antibiotic, antifungal, analgesic, antipyretic, antiviral, antiseptic, antidepressant, and antidiabetic.The dataset exhibited an approximately equal proportion of prescription and over-the-counter medications.

The prescription medicines that are most frequently dispensed are analgesic, antiseptic,antidepressant, antidiabetic, antiviral, antifungal and antibiotic. Analgesics given for have the highest count which means they are dispensed the most especially for viruses. Antivirals have the second highest count for being given for infections and wounds. The model correctly predicted the classification (prescription or OTC) for 49% of the instances in the test set. 49% of the instances predicted as prescription by the model were actually prescription medications. The model correctly identified 49% of the actual prescription medications.The F1-score, which balances precision and recall, is 0.49, indicating a relatively balanced performance between the two metrics.The current model exhibits moderate performance, with room for significant improvement.