

CHAPTER I

INTRODUCTION

1.1 Background

Currently, Acute Respiratory Infections (ARI) are still one of the major health problems globally [1]. ISPA is an acute inflammatory condition of the upper and lower respiratory tract caused by infection with microorganisms, such as bacteria or viruses, whether or not they involve lung tissue (parenchyma) [2]. The disease includes various forms of infection, with varying severity depending on the type of pathogen, the patient's body condition, as well as environmental factors [3].

ISPA is known to be the leading cause of morbidity and mortality due to infectious diseases worldwide, nearly 4 million people die from acute respiratory infections each year [4], and has a fairly high rate of transmission [5]. In Indonesia itself, ISPA is classified as a common disease experienced by children to adults, with the frequency of coughing accompanied by fever averaging 3 to 6 times per year. ISPA is also the main cause of the high number of visits to health facilities, which is around 40%–60% of the total visits to health centers and 15%–30% of outpatient and hospitalized visits [6].

After understanding the urgency of ISPA disease, the author collected data directly through a case study at RH Farma Surabaya Pharmacy in the period January 2025 to April 2026. The data obtained was in the form of drug prescriptions redeemed by the patient and information on the symptoms experienced. Based on this data, it is known that most people who redeem drugs experience symptoms related to ISPA, such as cough, fever, shortness of breath, or sore throat.

However, because ISPA covers various types of diseases with fairly broad characteristics, further grouping is carried out to identify the most dominant diseases. From the results of the analysis based on direct data through a case study at the RH Farma Surabaya Pharmacy in the January 2025 period, it is known that the most common symptoms lead to indications of bronchitis, which is one of the common types of ISPA [7]. Acute bronchitis itself is defined as inflammation of the respiratory tract, from the trachea to the bronchi, caused by an infection, and is generally characterized by a phlegm or dry cough that lasts less than three weeks [8]. In Indonesia, cases of bronchitis reach around 1.6 million people, and in East Java in particular, the prevalence is quite high, especially in children 16.3%, adolescents 56.5%, and adults 27.2% [9].

In disease prediction systems, machine learning algorithms have a very important role in analyzing data and generating accurate predictions. One of the algorithms that has been considered effective in this regard is CatBoost [10]. CatBoost has its own advantages through a gradient boosting mechanism that is specifically designed to handle heterogeneous data. This algorithm is able to improve model stability and provide optimal prediction performance. The approach used in CatBoost utilizes efficient coding techniques to minimize the risk of overfitting. This strategy contributes greatly to improving the overall accuracy of the model, as CatBoost was developed with the primary goal of addressing the weaknesses of conventional boosting algorithms [11]. In addition, the use of CatBoost in the health field is still relatively rare, when compared to other popular algorithms such as XGBoost or LightGBM, thus providing a fresher and more innovative scientific contribution [12].

Although CatBoost is capable of generating predictions, this algorithm is a black box because it does not directly explain the reason behind the predictive results given. To answer these challenges, this study applied a model interpretation method, namely SHAP (SHapley Additive exPlanations). In the context of prediction, the SHAP value represents the contribution of each feature to the model's output in a fair and measurable manner. Each feature exerts a different magnitude and direction (positive or negative), depending on its role in determining the predictive results produced by the model [13].

Using SHAP encapability, each feature is analyzed for its contribution to the output value produced. The results show whether a symptom drives the prediction in a positive (indicated bronchitis) or negative (non-bronchitis) direction. Thus, it can be identified which features have the most influence on the model's decisions. This analysis provides a more transparent and interpretive understanding of the algorithm's decision-making process.

The results of this interpretation are then used to design a drug recommendation system. By knowing the main symptoms that drive the prediction of bronchitis, the system can recommend the appropriate type of medication based on previous patient data patterns. To support consistency and clarity in recommendations, this system is designed using a rule-based system approach, which is by establishing rules based on the symptoms indicated by bronchitis. Thus, if in the future there are patients with similar symptom patterns, pharmacists can immediately provide appropriate drug recommendations without the need to go through repeated medical consultation processes, especially for mild cases that commonly occur.

In general, the developed system is expected to be able to encourage the improvement of the quality of health services at the primary level. The main focus is directed at improving work efficiency, accelerating the decision-making process, and accuracy in

providing appropriate interventions to patients. This predictive approach provides opportunities for healthcare workers, especially in pharmacies, to respond to patient complaints more quickly, precisely, and data-driven.

1.2. Problem Formulation

Based on the above background description, the problems that can be formulated are as follows:

1. How can the CatBoost algorithm be used to predict Bronchitis Disease based on patient symptom data at RH Farma Surabaya Pharmacy?
2. How can the SHAP approach be used to interpret the most influential symptom factors in the prediction of Bronchitis Disease using the CatBoost algorithm?
3. How can the combination of the CatBoost algorithm and the SHAP method improve accuracy and transparency in bronchitis prediction?
4. How can the results of prediction interpretation be used in a rule-based system of drug recommendations to support decision-making at RH Farma Pharmacy?

1.3. Purpose

Based on the formulation of the problem that has been determined, the objectives of this research are as follows:

1. Analyzing the application of the CatBoost algorithm in predicting Bronchitis Disease based on patient symptom data at RH Farma Surabaya Pharmacy for the period January 2025 to April 2026.
2. Using the SHAP method to identify the main factors of symptoms that contribute to the predictive outcome of Bronchitis Disease.
3. Evaluate the effectiveness of the combination of the CatBoost algorithm and the SHAP method in improving the accuracy and transparency of bronchitis disease prediction as the basis of a rule-based system of drug recommendations.

1.4. Benefits

This research is expected to provide benefits for various parties as follows:

1. For Academics and Researchers:
 - a. Adding insights into the application of CatBoost algorithms and SHAP methods in predictive analysis based on health data.
 - b. Provide a reference for further research in the field of machine learning, medical data analysis, and prediction of Bronchitis using patient symptom data.
2. For Pharmacies and Medical Personnel:

- a. Assist pharmacies in identifying the pattern of patients' symptoms associated with Bronchitis Disease to support early diagnosis.
3. For Patients and Society:
 - a. Improve the efficiency of health services with more practical recommendations based on symptom data.
 - b. Assist patients in getting more appropriate treatment according to the results of disease prediction based on the symptoms experienced.

1.5. Problem Limitations

In order for this research to be more focused and directed, the limitations of the problems set are as follows:

1. This study only used patient symptom data and drug purchase history at RH Farma Surabaya Pharmacy in a certain period.
2. The data used in this study were limited to patient symptom information and medication purchase history related to Bronchitis, so it did not include other medical data that may be relevant (e.g. laboratory test results, more in-depth history of other diseases).
3. The algorithm used in this study is CatBoost for disease prediction and the SHAP method for interpretation of prediction results, without comparing with other algorithms.
4. This study focused only on predictive analysis and interpretation of symptom factors, without direct clinical intervention.
5. The results of this study are intended to identify patterns and symptom factors related to bronchitis indications based on pharmacy data, and are not intended as an official medical diagnostic tool.