CHAPTER 1

INTRODUCTION

1.1 Background

Recently, the development of internet traffic has increased. Most of this traffic is in the form of content consisting of video, sound, text, or images. The condition of the current network architecture is still IP-based. This will become an obstacle in the future if the traffic requested by users gets more significant and more varied because IP-based networks will transmit continuously, which can result in the information being sent not being fully received. Named Data Networking (NDN) is a paradigm in content-based networks. In NDN networks, routers can perform caching or storage so that user requests do not have to come from servers [1]. On NDN, the user will send the interest as needed, and then the interest will be received by the router, which will then look to see whether the requested interest is available in the Content Store (CS) on the router. If the requested content is available, it will be sent back to the user, but if not, it will be forwarded to the next router based on the Forwarding Information Base (FIB). With such a model, it can increase the speed of content distribution and reduce the request time from users. Besides that, it can maximize the resources available on the network.

One of the most critical parts of NDN is caching, divided into two parts: caching decision policy and caching replacement policy. Caching decision policy is related to what content should be stored while caching replacement policy is related to where the content is placed. Caching is concerned with storage capacity and the effectiveness with which content is retained. Constraints encountered in caching include limited content storage size and storage that performs storage regardless of whether the content is popular, thereby reducing resources besides storing the same content on each router [2]. Research on caching related to popular content has been carried out a lot. One of the solutions associated with popularity is content prediction, such as the research conducted by Wei [3] predicting content popularity using the Deep-Learning-based Content Popularity Prediction (DLCPP) method on ICN can improve the accuracy of the content requested by the user. However, the designed architecture still requires parameter adjustments to get maximum results. Furthermore, research was conducted by Qi Chen [4] to improve the accuracy of content popularity predictions with user clusters that adopted the ARMA (Autoregressive Integrated Moving Average) model in their manufacturing.

Another method to optimize NDN caching can be implemented through proactive caching. Proactive caching is an approach that stores content before the user requests it. This approach will increase the user's efficiency in obtaining content. Siyang Shan [5] stated that proactive caching on a volatile network reduces the number of hops and improves performance. Research conducted by Zhe Zhang [6] said content popularity and proactive caching on ICN networks for autonomous vehicles could increase caching efficiency in terms of hit ratio and the number of hops.

Based on Zhe Zhang, research was carried out on a vehicular network with mobile user characteristics and a continuously changing topology. Implementing proactive caching can increase efficiency. As a result, the authors intend to investigate combining popular content prediction with proactive caching in NDN architecture to improve caching efficiency in terms of a cache hit ratio and round-trip time.

1.2 Problems Definition

Based on the context underlying this research, the following problem statement has been constructed:

1. How does implementing popular content prediction affect proactive caching in Named Data Networking?
2. What are popular content prediction models used in proactive caching in Named Data Networking?

1.3 Research Objective

This study aims to design a deep learning model for predicting the following popular content. This model will be applied to proactive caching on NDN to get a high cache hit ratio and low round trip time.

1.4 Scope of Works

1. The testing is conducted using an emulator and on a laboratory scale.
2. The machine learning model is used to determine the storage of content on NDN nodes.

1.5 Hypotheses

Implementing proactive caching based on deep learning models to predict popular content can reduce round trip time, and increase cache hit ratio on Named Data Networking (NDN).

1.6 Research Methodology

1. The first step is to do a literature review, which means looking for references to books and journals on NDN and ANN. The journal will serve as a resource for research that has already been conducted and potential future studies. While the book serves as a theoretical resource for research

2. Perform exploratory data analysis on the MovieLens dataset. In this step, the data will be cleaned and checked regarding the data type and its completeness. This stage is an essential part because it will affect the results of the predictions made.

3. The next step is to do feature engineering and feature selection because not all dataset features are used. One method of selecting features is by measuring univariate statistical values.

4. Create a deep learning model to predict popular content based on the MovieLens dataset using an Artificial Neural Network (ANN) algorithm.

5. valuation of the prediction model that has been made, if you get a model that is not optimal, you will do hyper-parameter tuning of the model.

6. Simulating NDN using mini-NDN based on data obtained from machine learning, content placed on NDN will be based on that model.

7. Analysing the simulations performed on mini-NDN to determine the effect of implementing provocative caching and predicting content popularity on the hit ratio and the round trip time on NDN.

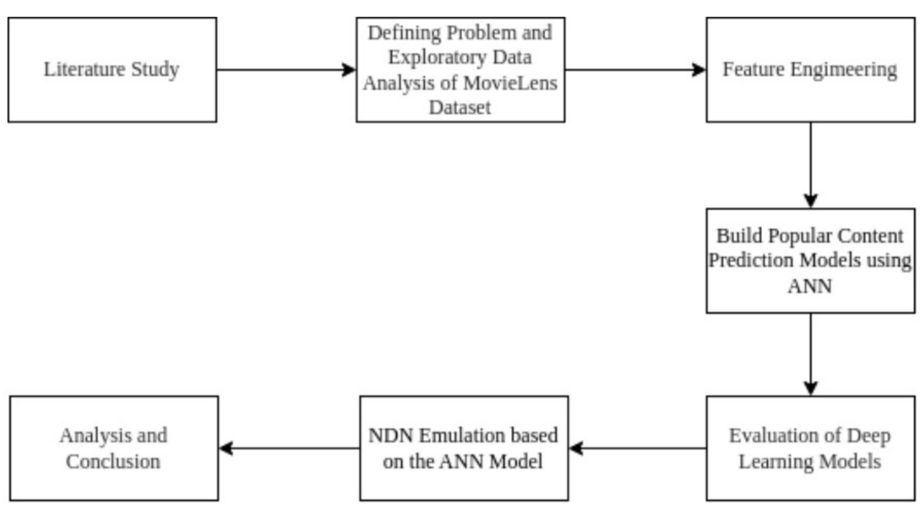

Figure 1.1: Research Methodology