Optimization of K-Nearest Neighbor for Recommendation System

BY

Nazmul Hassan
ID: 191-15-12282
AND

Nusrat Jahan Nira
ID: 191-15-12620

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Md. Abbas Ali Khan
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

Md. Ferdouse Ahmed Faysal
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This Project/internship titled **"Optimization of K-Nearest Neighbor for Recommendation System"**, submitted by Nazmul Hassan, ID No: 191-15-12282, and Nusrat Jahan Nira, ID No: 191-15-12620 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on January 23, 2023.

BOARD OF EXAMINERS

Chairman

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Md. Zahid Hasan
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Fahad Faisal
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Dr. Ahmed Wasif Reza
Associate Professor
Department of Computer Science and Engineering
East West University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Md. Abbas Ali Khan, Assistant Professor, Department of CSE** Daffodil International University and co-supervision of **Md. Ferdouse Ahmed Faysal, Lecturer, Department of CSE,** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Md. Abbas Ali Khan
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Faysal

Md. Ferdouse Ahmed Faysal
Lecturer
Department of CSE
Daffodil International University

Submitted by:

নাজমুল হাসান

Nazmul Hassan
ID: 191-15-12282
Department of CSE
Daffodil International University

Nusrat

Nusrat Jahan Nira
ID: 191-15-12620
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Md. Abbas Ali Khan**, **Assistant Professor** and **Md. Ferdouse Ahmed Faysal, Lecturer**, Department of CSE, Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, and valuable advice reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Touhid Bhuiyan**, Professor, and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

A tool for interacting with vast and complex information spaces is a recommender system. They offer a tailored view of these areas, giving users the items, they are most likely to find interesting priority. The field has advanced significantly since 1995 in terms of the range of issues it addresses, the methods used, and the applications it can be put to use. Research on recommender systems incorporates a variety of artificial intelligence methods, such as constraint satisfaction, case-based reasoning, user modeling, machine learning, and data mining. Many online e-commerce platforms, including Amazon.com, Netflix, Pandora, and others, heavily rely on personalized recommendations. This wealth of real-world application knowledge has propelled researchers to broaden the application of recommendation systems to new and difficult domains. Because there is such a high demand for both online shopping and movies, businesses rely on it. Machine learning-based technology makes it easier than ever to locate our true target audience. Jobs that aid in our understanding of our requirements and offer suggestions for them are recommended. Items that consumers are looking for. In this study, various machine learning algorithms for recommending various product purchases are compared.

Contents **PAGE**

Board of examiners……………………………………………………………………………..................ii

Declaration……………………………………………...……………………………………………...iii

Acknowledgements………………..…………………………………………………………………...iv

Abstracts…………………..……………………………………………………………………………v

CHAPTER **PAGE**

CHAPTER 1: INTRODUCTION 1-4

1.1 Introduction 1

1.2 Motivation 2

1.3 Rationale of the Study 2

1.4 Research question 3

1.5 Expected Outcome 3

1.6 Research Management 3

1.7 Lay Out of the Report 4

CHAPTER 2: BACKGROUND 5-8

2.1 Preliminaries and Terminologies 5

2.2 Related Works 6

2.3 Scope of the problem 8

2.4 Challenges 8

CHAPTER 3: RESEARCH METHODOLOGY 9-16

3.1 Introduction 9

3.2 Research Subject and instrumentation 9

3.3 Workflow 9

3.4 Data Collection Procedure 12

3.5 Proposed Methodology 12

3.6 Implementation Requirement 16

CHAPTER 4: EXPERIMENTAL RESULT AND DISCUSSION 17-18

4.1 Experimental Setup 17

4.2 Performance Analysis 17

4.3 Result Discussion 18

CHAPTER 5: IMPACT ON SOCIETY, ETHICAL ASPECTS AND SUSTAINABILITY 19-20

5.1 Impact on society..19

5.2 Ethical Aspect..20

5.3 Sustainability...20

CHAPTER 6: SUMMERY, CONCLUTION, RECOMMENDATION, IMPLICATION FOR FUTURE RESEARCH 21-22

6.1 Introduction 21

6.2 Conclusion...21

6.3 Future Work 22

References 23

APPENDICES...25

PLAGIARISM REPORT 26

LIST OF FIGURES **PAGE**

Figure 3.5: Workflow Diagram……………………………………………………………………………..11

Figure 4.2.2: R2 score of Algorithm…………………………………………………………………..18

LIST OF TABLES **PAGE**

Table 4.2: List of Algorithms……………………………………………………………………………….17

CHAPTER 1

Introduction

1.1 Introduction

To recommend products that users are likely to be interested in, recommendation systems try to predict user preferences. These are among the most effective machine learning methods used by internet retailers to increase sales. The explicit user ratings that follow watching movies or listening to music, implicit search engine queries or purchase history, or any other information about the user/item itself are the sources of the data needed for recommender systems. These data are used by platforms like Spotify, YouTube, or Netflix to recommend playlists, so-called Daily mixes, or to recommend videos, respectively. With the growth of YouTube, Amazon, Netflix, and many other such internet services over the last couple of decades, recommender structures have occupied a larger and larger portion of our lives. Recommender systems are now inescapable in our daily online activities, whether it be in e-commerce (proposing to customers items that would interest them) or online advertising (proposing to customers the right content, matching their preferences). Generally speaking, recommender systems are algorithms that make relevant recommendations to users for things like movies to watch, books to read, goods to buy, or other things depending on the industry. In some industries, recommender systems are unquestionably important because they have the potential to make sizable profits while still being environmentally friendly or to help businesses stand out significantly from their rivals. We can cite the fact that, a few years ago, Netflix organized a competition (the "Netflix prize") with the goal of supplying a recommender device that plays above its very own set of rules and a prize of $1 million to win as evidence of the importance of recommender structures.

1.2 Motivation

We want to look back at the development of e-commerce in order to understand the motivation behind recommender systems. The majority of goods were sold in brick-and-mortar stores prior to the growth of e-commerce. The inventory of a store could only be as large as the actual store, and goods that didn't sell well were unprofitable. Due to a rigid inventory, retailers were forced to sell only the top products of contemporary thought. The development of the online marketplace in the middle of the 1990s completely changed the retail sector. Because this new kind of digital marketplace permitted unlimited inventory, vendors could broaden their product offerings to include more specialized, less mainstream items. According to Chris Anderson in The Long Tail, "the mass market is popping into a mass of niches.". Niche products are able to outsell bestsellers by overcoming inventory restrictions. Due to this, niche products are now more expensive than they were previously for e-commerce companies. People are less likely to buy when they have too many options. Beyond just grocery shopping, having too many options can influence people's online retail decisions. By limiting the search space and locating the highest-quality items that are most pertinent to users, recommender systems lessen selection overload.

1.3 Rationale of the Study

There is no doubt that hundreds of works have been done in the field of recommendation systems using the KNN algorithm However, just a few works on optimizing KNN with the Hyperparameter algorithm. Our work is a novel technique for a recommendation system that uses a fully upgraded dataset which will produce even better results than earlier efforts and will serve as a powerful dataset for future recommendation systems.

1.4 Research question

Defining a clean, concise, and centered studies query is an essential first step in starting studies work. It identifies exactly what we need to discover and gives a clean recognition and purpose. To reach a realistic, efficient, and correct technique for this problem, the researchers would want to provide subsequent inquiries to speak their minds and findings.

- Can we implement an optimized KNN for our recommendation system research?
- Can we preprocess the raw data abstract from the data source?
- Can this work improve the field of recommendation systems?

1.5 Expected Outcome

In this section, we have mentioned the proposal of our research based on the research question. We expect to build a prominent model for the recommendation. The researchers would like to propose the following expected outcome of the research.

- Successfully implemented an optimized KNN for the recommendation system.
- This work can improve both businessmen and consumers.

1.6 Research Management

At the end of the research project, we first created the planned architecture and implemented it step by step. Through project management until the goal is achieved success. The project flow is shown below.

I. Conducting logistics searches for research related to our project.

II. Selected research studies relevant to our research themes.

III. Conduct analytical studies on all selected studies to learn more about the recommendation system.

IV. Collect specific records from the recommender system database.

V. Preprocess and filter datasets for use in machine learning classification algorithms.

VI. To achieve our goal, we have implemented the most effective deep learning classifier.

VII. We have trained and tested our classifier to show how accurate it is at work.

VIII. Documented completed research projects.

1.7 Lay Out of the Report

Chapter 1 covers an introduction to the study including the motivation for the study, research rationale, research questions, expected results, project management and funding, and the overall structure of the project.

Chapter 2 describes the background of the research. preliminary work, related work, Comparative analysis and summary, problem scope, and challenges.

Chapter 3 provides a theoretical analysis of the study. The first part of this chapter Shows the progress of the project. Next, we describe the data collection process. Furthermore, this chapter presents an algorithmic approach to a machine learning classifier. Finally, we will discuss some implementation requirements.

Chapter 4 includes experimental results, a discussion of results, and project performance. Contains experimental visualizations to aid in the understanding result.

Chapter 5 also discusses the impact of research on society and the environment also some ethical issues.

The last chapter, Chapter 6, contains the conclusion and summary of the work. The last chapter section highlights some of the chapter's shortcomings and eliminates some recommendations for future research.

CHAPTER 2

Background

2.1 Preliminaries and Terminologies

The purpose of a recommender system is to provide users with recommendations based on a number of different criteria. These algorithms foretell the goods that customers will be most interested in and likely to buy. Recommendation systems are employed by businesses like Netflix and Amazon to guide customers toward appropriate books and movies. Recommender systems analyze a large amount of available data by selecting the most pertinent information based on user-provided information and other variables that take into account user preferences and interests. They also look for matches between users and items and attribute similarities between users and items for recommendations. A system like this is advantageous to users and the services offered. These systems have also enhanced quality and judgment.

1. Popularity- Based Recommendation system:

It is a type of recommendation system that operates according to trends and/or popular theory. With the help of these systems, users can receive direct recommendations for the hottest products, TV shows, and movies. For instance, the system will recognize that a product is the most popular if it is frequently purchased by the majority of people. Therefore, the system will also recommend this product to all newly registered users. As it becomes more expensive, more people buy it.

2. Content-Based Recommendation system

Another recommendation system that bases its decisions on content similarity is this one. When a user watches a movie, the system looks for other films in the same genre or with similar content as the one they are currently watching. In order to determine similarity when looking for related content, some fundamental attributes are used.

3. Collaborative Filtering:

It is regarded as one of the most intelligent recommender systems that take into account user similarities as well as frequently used resources like e-commerce and streaming video websites. Check similar users' preferences and offer suggestions. The degree of similarity can refer to both the degree of similarity between various objects as well as user preferences. When there is a wealth of data about users and items, the system makes recommendations that are more effective. [1]

2.2 Related Works

Over the years a number of research works have been done on recommendation systems by implementing the KNN algorithm. In 2019, a research paper titled "Book Recommendation System Using KNN Algorithm" was published, in which they used amazon.com's dataset by collecting user reviews and applied the KNN algorithm to come up with a better Publish book recommendation system. Various techniques have been introduced to recommend items such as content, collaborative techniques, and association mining. This project helps solve the sparse data problem by using the KNN algorithm and association rule mining to achieve better performance**.** [2] In 2020, a machine learning algorithm based titled "Book Recommendation System using Machine Learning" was released. In this research, they used machine learning algorithms, K-NN, and matrix factorization. Cooperative filtering first collects ratings or likes for books provided by multiple users and then suggests books for different people based on different past likes and likes. [3] In 2020, a study, on "Movie Recommendation System with Collaborative Filtering Using K-NN", in which collaborative Filtering is used, scores data in a matrix format that represents movies or videos in columns and users in rows to implement the K-Nearest Neighbors model measures user Similarity using a distance metric in the gaps for an unassigned grade and then make a recommendation. [4] In January 2019, an article titled

"Movie Recommender System Using K-Means Clustering AND K-Nearest Neighbor" was published. The proposed work focuses on introducing various concepts related to machine learning and recommendation systems. Various tools and techniques have been used in this work to create recommender systems. Various algorithms like K-Means Clustering, ANN, and Collaborative Filtering. Collaborative filtering first collects movie ratings or movie likes from different users and then recommends movies to different users based on past similar interests and hobbies. Another supporting technique used in developing a recommender system is clustering. Clustering is a process of grouping a set of objects together so that the objects in the same cluster are more similar than the other cluster. [5] In the past 2017, "Design and implementation of a movie recommendation system" based on the collaborative KNN filter algorithm", where the main research content is to help users automatically include movies of their interest in the big data of movie information using the KNN algorithm and collaborative filtering algorithm and developed a prototype movie recommendation system based on KNN's collaborative filtering algorithm. [6] Recently published in 2021, an article titled "Music Recommendation System with Advanced Classification "describes a collaborative recommendation algorithm that uses filtering and content filtering to combine network output with log files to recommend music to users in a personalized music recommendation system. The proposed system includes a log file that stores the previous history of a user's music playlist. The suggested music recommendation system provides music recommendations for each recommendation from a content-based approach that extracts user history from log files and makes. [7] In a research paper entitled "Applied Parallel Computing K-Nearest Neighborhood-based Music Recommendation System," They chose the K-Nearest Neighborhood (K-NN) model to predict song ratings. This model is an element-based algorithm that finds neighborhoods between elements (songs in this context), in contrast to user-based algorithms that find neighborhoods between users. Her goal for this project is to improve prediction results using the KNN algorithm. It reduces the total run time of the parallelism. [8] In 2020 they published "PRODUCT RECOMMENDATION SYSTEM USING MACHINE LEARNING TECHNIQUES" in which they created a matrix to represent the ratings that the various users have given to the product. They then created the user profile and article profile and make recommendations based on similarity to the other groups. Next, we'll apply the N-algorithm, which is part of machine learning, and Pearson's correlation coefficient to give recommendations to the user. To assess the performance of the model, RMSE was used, which provides information about the closeness of the generated recommendations. [9] Another related work in 2019 titled "Location-Based Farm Products Recommendation System Using Novel KNN Algorithm," uses an improved Novel KNN algorithm to find the closest vendor K by measuring the distance between the sellers and buyer computed using a Euclidean and distance metric. Details Posted by farmers and buyers are stored in a database and updated dynamically. The recommendation system recommends the closest sellers and their products based on buyer interest. System performance is analyzed in terms of accuracy and mean absolute error. [10]

2.3 Scope of the problem

We set some boundaries and parameters within which our research work had been conducted. These boundaries refer to the scope of the study and define all the variables that were considered in our research project, as well as the parameters upon which the study operates. As per the scope of research, our study explored the recommendation. We worked with the R2 score value.

2.4 Challenges

The primary challenge of the research work for us was to collect datasets and choose algorithms. When we choose the algorithm, we found that it was normal KNN. They have to find the best way for optimized KNN. We used hyperparameters to optimize the KNN algorithm. We faced the challenge of finding out the proper and necessary features. We overcame that challenge by cleaning and pre-processing the raw dataset and only used the necessary features for our prediction model cleaning and pre-processing the raw dataset and only used the necessary features for our prediction model.

CHAPTER 3

Research Methodology

3.1 Introduction

In this portion, we summarize the workflow of our proposed model to classify recommendation systems. Core points such as data collection, processing, and proposed models are also expressed with equations, graphs, tables, and descriptions. An optimized KNN has been applied to our dataset which we have collected from Kaggle and applied this dataset in the KNN algorithm with Hyperparameter Optimization for better performance. The last part of this chapter explains the statistical theory of our project and we also provide the implementation requirement which we needed.

3.2 Research Subject and instrumentation

Research topics include topics that help us understand or clarify topic concepts. Implementation of design models, collection of datasets, Processing the dataset, training the model, and making changes based on its record. Another section, Instrumentation, is basically about the techniques and methods used. In this proposed work, we used the Windows platform as programming. NumPy, Pandas, Skit Learn, Matplotlib, Seaborn, etc. Google Colab is used throughout the training and testing process, Google Colab is a web-based IDE for Python and can also use cloud storage for proposed work applications.

3.3 Workflow

This portion will describe the workflow which we have executed to go through our project. A few stages have been executed such as data collection, data processing, model selection, and lastly evaluating the outcome and future scope of this problem.

Stage 1- Data Collection: We have collected the raw data from online platforms such as Kaggle. This site contains different datasets for different issues, especially different types of recommendation systems such as books, movies, music, products, news, etc. the machine reads the data set from the device. These data sets are used by the machine as training sets to complete the project. Using these datasets, recommendations are made. Among them, we choose Book Recommendation System.

Stage 2- Data processing: We don't need to preprocess the data. We directly input the raw data into the machine.

Stage 3- Model Selection: Among various types of ML algorithms, we have chosen the KNN and Hyperparameter Optimization (HPO) algorithms. HPO algorithms than have a subset of this algorithm such as Grid Search, Random Search, BO-GP, and Genetic Algorithm. We used this optimized algorithm with KNN to perform better outcomes. Lastly, Optimize KNN outperformed the recommendation score of every class.

Stage 4- Performance Evaluation: This section analyzed the performance from the proposed model on the dataset which contains a few training and testing accuracy and loss graphs. The final section shows the output of predicted classes.

Stage 5- Conclusion and future work: This portion would include the future scope of our project and shortly describes the whole process.

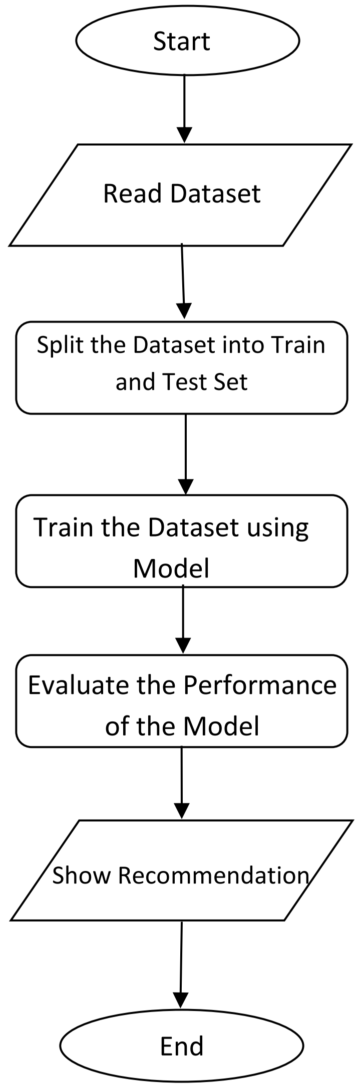

Figure 3.3: Workflow Diagram

3.4 Data Collection Procedure

The data collection phase of the implementation process is the first term. The appropriate dataset is selected at this time to enable further computations. The dataset for the recommendation system is taken from the Kaggle website. We collected Book Recommendation System from kaggle where information about books and user ratings are included. User ratings on a scale of 1 to 5 make up the dataset. Additional calculations are made using this data and the Python programming language.

3.5 Proposed Methodology

3.5.1 KNN

The K-Nearest Neighbors (KNN) computation is characterized as a distance-based supervised learning computation and is therefore the only computation used to solve classification problems. According to this algorithm, we first use the classes of the k nearest neighbors to allocate unused foci and assign them to lessons that can also accommodate most of the k neighbors. Various distance measures can be used to find the nearest neighbors. The most commonly used method is to determine the "Euclidean distance". Moreover, this computation is a non-parametric, agnostic learning computation. The rationale for the ANN algorithm is to use centroids of information separated into several classes in the database to predict the classification of the most recent test points. [11]

3.5.2 Hyperparameter optimization

One of the biggest obstacles to implementing machine learning solutions is model optimization. Model optimization has been the focus of entire branches of machine learning and deep learning theory. In machine learning, hyperparameter optimization seeks to identify those of an algorithm's hyperparameters that produce the best-measured performance on the validation set. In contrast to model parameters, hyperparameters are

set by machine learning engineers before training. A random forest's tree count is a hyperparameter, whereas a neural network's weights are model parameters that are learned during training. In my mind, a model's hyperparameters are the settings that need to be adjusted in order for it to best address a machine learning challenge.

- How quickly a neural network learns new information.
- The C and support vector machine hyperparameters.
- The letter K in "k-nearest neighbors. [12]

3.5.3 Grid Search

Despite being the simplest to use and comprehend, the grid search method is ineffective when the number of parameters is high and loosely constrained under H0. Let * be the set of perturbation parameters that maximize the p-value, where * = (1, 2,..., m). Assigning a vector of lower bounds (a) = (a1, a2,..., am) and upper bounds (b) = (b1, b2,..., bm) to each component makes setting up a lattice search a simple process. is. definition. Every interval of the form [ai, bi] in a grid search contains the values ai and bi, and there are n points that are evenly spaced apart. This causes nm grid points to be checked in total. The highest value between each pair of points is then selected after the calculations are complete. [13]

3.5.4 Random Search

In the random search method, the model is trained using a set of hyperparameters that are selected at random. We choose the most effective set of random hyperparameters. Grid search and random search are somewhat similar. The main distinction is that we don't offer a range of potential values for each hyperparameter. Rather, we use each hyperparameter's statistical distribution to determine its values. For each hyperparameter, a sampling distribution is set up so that a random search can be done. We can manage the number of hyperparameter combinations tested using this technique. In contrast to grid search, which tries every combination, random search lets we choose how many models to train. Iterations of a search can be based on computational power or the amount of time required for each iteration. [14]

3.5.5 Bayesian Optimization with Gaussian process (BO-GP)

A model-based black-box optimization algorithm called Bayesian optimization is tuned for a very expensive objective function, also known as a cost function. Bayesian optimization is a black-box optimization algorithm that looks for the highest possible value of an unknown objective function that can be sampled (for instance, by evaluating a robot's performance). Similar to every model-based optimization algorithm (including kriging, DACE, and surrogate-based algorithms), Using regression techniques to build a model of the objective function, Bayesian optimization then uses that model to choose a capture point and model the data. Since the general formulation of this algorithm computes the posterior distribution of the objective function using the probabilities of the previously collected data and the prior probabilities of the function type, it is known as a Bayesian algorithm.

A popular method for Bayesian optimization, Gaussian process regression, is used by Limbo to find models. Due to the way that they model both the cost function and the uncertainty surrounding each prediction, gaussian processes are particularly intriguing for regression. A Gaussian process establishes a probability distribution of potential values for a normally unknown cost function f(x) for each point x. Due to the Gaussian nature of these probability distributions, their mean and standard deviation can be used to describe them.

On the contrary, for every x, and can vary. So, we specify the function's probability distribution.

$P(f(x)|x) = N(\mu(x), \sigma2(x))$

The standard normal distribution N is represented here. We must apply a Gaussian process to the data in order to calculate (x) and (x). Assume for this purpose that each observation f() is a sample from a normal distribution. If we have a collection of data derived from a

number of observations, such as f(1,2), The vector is [f(1),f(2), for f(t). ,f (t)] is a sample from the multivariate normal distribution represented by the mean vector and covariance matrix. In this way, the n-variate normal distribution is generalized by the Gaussian process. n is the total number of observations. One observation is linked to another by a covariance matrix.

Due to the function's propensity to be smoother and the function built into the algorithm's function having a higher priority on the probability of insertion, two observations corresponding to nearby 1 and 2 values may be correlated. is a presumption that is made based on the fact that there are the two observations that correspond to the remote values of, 1 and 2 should not have an impact on one another (their distributions should be uncorrelated). In other words, the covariance matrix demonstrates that samples that are close to one another are strongly correlated while samples that are far apart are largely uncorrelated. A kernel function called k(1, 2) defines this covariance matrix. The usual foundation for this is the Euclidean separation between 1 and 2. The Gaussian process is determined as, given a set of observations P1:t=f(1:t) and the sampling noise 2noise (a user-defined parameter).

$P(f(x)|P1:t, x) = N(\mu t(x), \sigma 2t(x))$

where:

$\mu t(x)=k^{\intercal}K-1P1:t$ [15]

3.5.6 Genetic Algorithm

In order to solve optimization issues in machine learning, a genetic algorithm is a research-based algorithm. The importance of this algorithm lies in the fact that it quickly and efficiently resolves challenging issues. Numerous real-world uses for it exist, including data centers, electronic circuit design, decoding, image processing, and artificial life. [16]

3.6 Implementation Requirement

A list of prerequisites for such a mutation classification project has been produced after a comprehensive analysis of the required statistical or theoretical concepts and procedures. The components listed below are almost certainly going to be required:

1. Compatible Hardware/Software Environment
2. Operating System (Windows 7 or above/Linux distro)
3. Random Access Memory (Minimum 4 GB)
4. Hard Drive

Model Developing Tools

1. Python Environment
2. PyCharm
3. Google Collaboratory
4. Jupyter Notebook

CHAPTER 4
Experimental Result and Discussion

4.1 Experimental Setup

In this chapter, the result of the study is presented and discussed with a classification report. The result and comparison of KNN and hyperparameter optimization by R2 score will be presented throughout this chapter.

4.2 Performance Analysis

Python libraries import: NumPy, Pandas, MAtplotlib, and sklearn. The dataset contains three excel files. The dataset of 90,000 users includes 1.1 million reviews of Movies.

To calculate the accuracy of our predictions available to use an ordinary statistical metric named R2 score. The R2 score is a very important metric that is used to evaluate the performance of a regression-based machine learning model. It is pronounced as R squared and is also known as the coefficient of determination. It works by measuring the amount of variance in the predictions explained by the dataset. Simply put, it is the difference between the samples in the dataset and the predictions made by the model.

Algorithms	Value
KNN	0.0507
KNN+Grid Search	0.0675
KNN+Random Search	0.0688
KNN+Bayes Search	0.0332
KNN+GP	0.0687
KNN+GA	0.0664

Figure 4.2.1: List of Algorithms

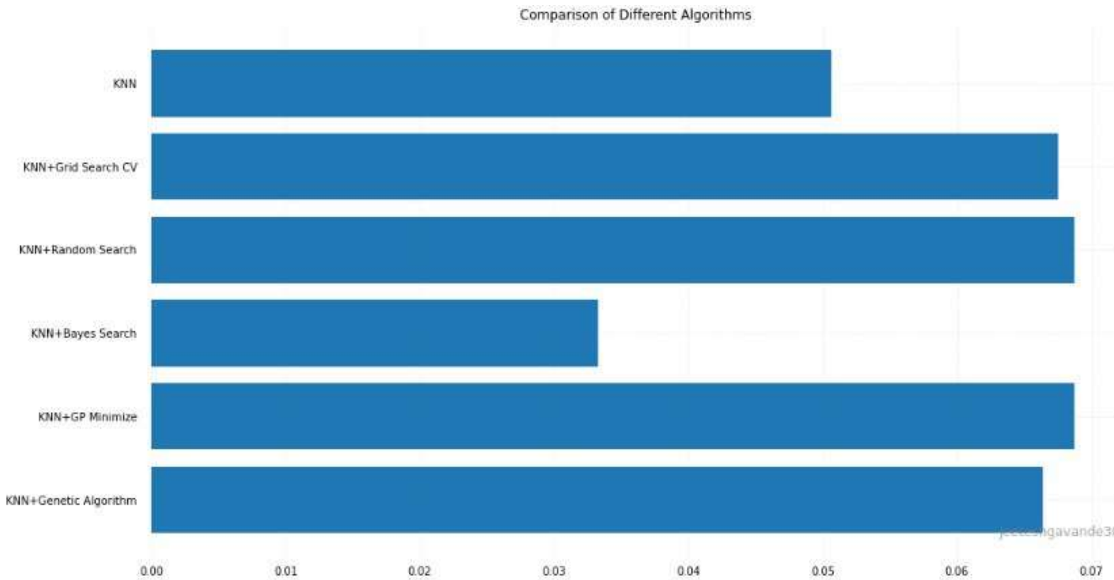

Figure 4.2.2: R2 score of Algorithms

4.3 Result Discussion

If we check our data table then we will find that our KNN value is 0.0507. In that case, our proposed algorithm's R2 score is more efficient, especially for KNN+Bayes Search Optimization.

CHAPTER 5

Impact on Society, Ethical Aspects, and Sustainability

5.1 Impact on society

Producing reports is not the real goal of these system operations. It is used to find flaws so that they can be fixed. We couldn't figure out why because Using our evaluation platform as a plug-in, are we able to create a recommendation system that can suggest logical next steps for improvement. Although they are still in the early stages of development, predictive analytics and recommendation systems are already a cornerstone of commercial e-commerce. After buying a rug, Amazon of course suggests a stool. After watching The Conversation, Netflix suggests watching The French Connection. What about logical recommendations to enhance the outcomes of classroom teachers' professional development activities?

E-commerce recommendation algorithms frequently use grouping frequencies, perhaps after taking some context and demographic information into account. Based on user and article usage trends, collaborative filtering offers a practical method for locating popular candidates. Though it may be a gauge of preference in social systems, popularity does not always point in the right direction. I don't think it's overstating things when I say that social advancement seldom goes in the direction of fashion. The availability of historical data is typically limited, which presents another difficulty in formulating recommendations in frameworks for welfare system improvement. Although there may be a wealth of information available globally, experts in this field also have anecdotal and systematically gathered information about what might be effective in a particular circumstance. It's unlikely to have a broad historical pattern of evaluation, recommendations, actions, and results before an evaluation within the system's own database. [17]

5.2 Ethical Aspect

We have protected the ethical facets of our research in accordance with the social responsibility principles. within our project. We used credible information from validated sources for our study. We have ensured the openness of scientific research through our work. In accordance with their ethical standards based on research, no type of business is harmed by our work. The entire working process is over. Two renowned research authorities closely monitor its execution.

5.3 Sustainability

Plan Since our research is visionary, it can ensure for a very long time on its own. The project can be completed. success in terms of every aspect of sustainability, including the environment and financial sustainability. Sustainability and community sustainability. Long-lasting models have been created. Because the project's resources are not easily lost. simple to maintain. To remain sustainable in the future, stay away from excessive complexity. The field of research is one that has potential for development. The Growth Aspect is Recognized. Because it is our project, we can continue to work on it in the future to keep it going.

CHAPTER 6

Summary, Conclusion, Recommendation, Implication for Future Research

6.1 Introduction

The recommendation right away in the field of e-commerce has been broadly used, but with the advent of the era of big data the original recommendation algorithm has been a great challenge. This paper details the recommendation system, while the K-neighbor algorithm is proposed for improvement. The improved algorithm improves the recommendation accuracy while reducing the calculation K- nearest neighbor on the cost. However, with the development of the information age, the explosive growth of information, and the large user of project information to make collaborative filtering technology is facing a great challenge, which will be the next focus of our research.

6.2 Conclusion

In this study, we proposed a machine, learning-based model. Product recommendations from a classifier. From the recommendation system resource database, we gathered the necessary datasets. To create the best dataset for our study, we collected raw data next when Utilizing the hyperparameter optimization algorithm and the KNN machine learning algorithm, and created the work. We followed the model developed and examined the outcomes demonstrating the goal's success. A lens with a high KNN and HPO recommendation for accuracy.

6.3 Future Work

Giving recommendations to the user in an e-commerce setting is the project's main objective using machine learning algorithms for a website. This was created by us and put into action. K-nearest Neighbor and Hyperparameter Optimization System are considered as the dataset. Based on similarity, ratings are left by other customers for a particular product. We attempt to recommend the products to our current users from among the rated products. The forthcoming work. Among the project's goals is to increase the system's effectiveness. Additionally, it ought to be able to give. Appropriate suggestions to users or to those who have never made a purchase before new clients. Future experiments could involve deep learning and recurrent neural networks. with the aid. We can get around some of the problems with matrix factorization by using deep learning techniques. Recurrent neural networks are used in deep learning to account for time in the recommender. A system that the matrix factorization method cannot achieve. We can also work on supplying subpar user recommendations, record the user's response and use it in the system.

References

[1] [Online]. Available: https://www.analyticssteps.com/blogs/what-are-recommendation-systems-machine-learning.

[2] Y. N. a. P. K. Bhagirathi, "Book Recommendation System using KNN Algorithm," *International Journal of Research in Engineering,* pp. 336-339, 2019.

[3] F. Ijaz, "Book Recommendation System using Machine learning," *EasyChair Preprint,* pp. 1-5, 2020.

[4] D. a. M. B. D. A. Bose, "Movie recommendation system with Collaborative Filtering using K-NN," pp. 1-4, 2020.

[5] R. A. S. a. A. N. Ahuja, "Movie Recommender System Using K-Means Clustering AND K-Nearest Neighbor," pp. 263-268, 2019.

[6] B.-B. Cui, "Design and Implementation of Movie Recommendation System Based on Knn Collaborative Filtering Algorithm," vol. 12, 2017.

[7] R. A. e. a. CH, "Music Recommendation System With Advanced Classification," 2021.

[8] S. a. V. G. Narayanan, "Applied Parallel Computing K-Nearest Neighborhood based Music Recommendation System".

[9] M. Kanagala, "Product Recommendation System Using Machine Learning Techniques," 2020.

[10] J. e. a. Sachin, "Location Based Agricultural Product Recommendation System Using Novel KNN Algorithm," 2019.

[11] [Online]. Available: https://en.m.wikipedia.org/wiki/K-nearest_neighbors_algorithm.

[12] [Online]. Available: https://www.kdnuggets.com/2020/05/hyperparameter-optimization-machine-learning-models.html.

[13] [Online]. Available: https://www.section.io/engineering-education/grid-search/.

[14] [Online]. Available: https://www.section.io/engineering-education/random-search-hyperparameters/.

[15] [Online]. Available: http://krasserm.github.io/2018/03/21/bayesian-optimization/.

[16] [Online]. Available: https://towardsdatascience.com/introduction-to-genetic-algorithms-including-example-code-e396e98d8bf3.

[17] [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-27609-5_17.

[18] [Online]. Available: https://medium.com/criteo-engineering/hyper-parameter-optimization-algorithms-2fe447525903.

[19] [Online]. Available: https://maelfabien.github.io/machinelearning/HyperOpt/#random-forests.

[20] [Online]. Available: https://www.analyticsvidhya.com/blog/2021/10/an-introduction-to-particle-swarm-optimization-algorithm/.

APPENDICES

While completing the research project, we encountered numerous difficulties. The first was determining the project's methodological approach. But the biggest difficulty we encountered during the research was that, as students with a background in computer science, we lacked research understanding regarding the Recommendation system. We overcame this obstacle by studying and analyzing related and earlier efforts in that sector. Our next task was to create the ideal dataset for our research to work on. The raw dataset that we got from the kaggle.

PLAGIARISM REPORT

Optimization of K-Nearest Neighbor for Recommendation System

ORIGINALITY REPORT

18%	12%	3%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
2	download.atlantis-press.com Internet Source	2%
3	hal.inria.fr Internet Source	2%
4	thecleverprogrammer.com Internet Source	1%
5	easychair.org Internet Source	1%
6	Submitted to Softwarica College Of IT & E-Commerce Student Paper	1%
7	www.ijert.org Internet Source	1%
8	Submitted to University of Sunderland Student Paper	1%
	www.nature.com	