# VISVESVARAYA TECHNOLOGICAL UNIVERSITY

**"Jnana Sangama", Belagavi-590018**


## INTERNSHIP REPORT [18CSI85]

**ON**

## "MACHINE LEARNING"

*Submitted in partial fulfillment of the requirements for the award of the degree of*

**BACHELOR OF ENGINEERING**

IN

**COMPUTER SCIENCE AND ENGINEERING**


Submitted by

**R. DEKSHITHA**

**USN: 1KS18CS075**


*Internship Carried Out at*

**Prinston Smart Engineers**

Vishwapriya Layout,Vishnuvardhan Statue Road
Begur Road
Bengaluru, Karnataka 560068


**Internal Guide**

**Mr. Prashanth H S**

**Assistant Professor**

**K.S Institute of Technology**

**Bengaluru-560109**

**External Guide**

**Mrs. Farheen Farhath**

**Co-Founder**

**Prinston Smart Engineers**

**Bengaluru-560068**



**Department of Computer Science & Engineering**

**K.S. INSTITUTE OF TECHNOLOGY**

No.14, Raghuvanahalli, Kanakapura Road, Bengaluru- 560109


**2021-202**

# K.S. INSTITUTE OF TECHNOLOGY

**BENGALURU-560109**

**Department of Computer Science & Engineering**

## CERTIFICATE

This is to certify that the Internship Training [18CSI85] entitled **"Machine Learning"** presented by **R.Dekshitha, USN: 1KS18CS075 of VIII semester** in partial fulfillment of the award of Bachelor of Engineering in CSE in Visvesvaraya Technological University, Belagavi during the academic year **2021-2022**. The Internship Report has been approved as it satisfies the academic requirements in respect of Internship Training work prescribed for the Bachelor of Engineering degree.

| | | |
|---|---|---|
| ---------- | ---------- | ---------- |
| **Signature of the Guide** | **Signature of the H.O.D.** | **Signature of the Principal** |
| **Prashanth H S** | **Dr. Rekha.B.Venkatapur** | **Dr. Dilip Kumar K** |
| **Assistant Professor,** | **Professor, Head of the Dept.,** | **Principal/Director,** |
| **K.S.I.T., Bengaluru** | **K.S.I.T., Bengaluru** | **K.S.I.T., Bengaluru** |

## External Viva

| **Name of the Examiners** | **Signature with Date** |
|---|---|
| **1.** | |
| **2.** | |

# COMPANY CERTIFICATE

PRINSTON SMART ENGINEERS

Course4job.com

Empowering and Inspiring future professionals

## Certificate of Internship

Machine Learning

R. DEKSHITHA

K.S INSTITUTE OF TECHNOLOGY

1KS18CS075

Intern Has Successfully Completed The Internship.
Intern Was Found To Be Good And Disciplined.
August 15th to September 20th, 2021.

CERTIFIED

PSEML2255

www.prinstonsmart.com

# DECLARATION

I, **R. Dekshitha** bearing the **USN: 1KS18SCS075** studying in the 8th semester of Bachelor of Engineering in Computer Science and Engineering at K S Institute of Technology, Bengaluru, hereby declare that the Internship Report entitled "Machine Learning", is a record of original work based on the Internship carried out at Prinston Smart Engineers. Under the External Guidance of **Mrs. Farheen Farhath**, Co-Founder, Prinston Smart Engineers and Internal Guidance of **Mr. Prashanth H S**, Assistant Professor, Computer Science & Engineering, K. S. Institute of Technology. The Internship Report has been submitted in partial fulfilment of the requirements for the award of the degree of Bachelor of Engineering in Computer Science and Engineering. The results embodied in this report have not been submitted to any other University or Institute for the award of any degree.

**Place**: Bengaluru **R. Dekshitha**

**Date**: **1KS18CS075**

# EXECUTIVE SUMMARY

The internship training at Prinston Smart Engineers was a storehouse of knowledge on Machine Learning. The university provided us an opportunity to work in any reputed organization as a part of the curriculum to get hands on experience of practical issues in real time situations. I went to PRINSTON SMART ENGINEERS as I had an interest in Machine Learning. As a part of training, I was introduced to various departments of the company and meeting great resource persons of the organization. I was allotted Mrs. Farheen Farhath as my trainer and I was following all the orders made by her. The guide made me more comfortable and the environment was student friendly. All the technical and non-technical staff was very helpful and very cooperative.

I was exposed to various tasks and was made to learn all the concepts and made me a good learner. The internship program consisted of learning all the technical activities of the organization. I learnt all the basics python and machine learning algorithms. All the present and olden techniques involved in the organization were made to learn in the internship program.

Finally, the company people let me to get a real time experience on working and the maintenance of the network. All the people in the organization helped me to successfully complete my internship training.

# ACKNOWLEDGEMENT

The successful presentation of the internship would be incomplete without the mention of the people who made it possible and whose constant guidance crowned my effort with success.

I take this opportunity to express my sincere gratitude to our **Management** K S Institute of Technology, Bengaluru for providing the environment to present the seminar.

I would express my gratitude to **Dr. K. V. A. Balaji,** C.E.O. K.S. Institute of Technology, Bengaluru, for facilitating me to present the seminar.

I would like to extend my gratitude to **Dr. Dilip Kumar K.,** Principal/Director, K.S. Institute of Technology, Bengaluru, for providing opportunity to present the seminar.

I thank **Dr. Rekha B. Venkatapur**, Professor and Head, Department of Computer Science and Engineering, K.S. Institute of Technology, Bengaluru, for her encouragement.

I whole heartedly thank Internship Coordinators, **Mr. Prashanth H S,** Assistant Professor and **Mrs. Beena K**, Assistant Professor, Department of Computer Science and Engineering, K.S. Institute of Technology, Bengaluru, for their support and guidance.

I would also like to thank my Internship guide Mr. Prashanth H S, Assistant Professor, Department of Computer Science and Engineering, K.S. Institute of Technology, Bengaluru, for his constant support.

Finally, I would like to thank all the teaching and non-teaching staff of the college for their co-operation. Moreover, I thank all my **family** and **friends** for their invaluable support and cooperation.

**R. Dekshitha**
**1KS18CS075**

# ABSTRACT

The Machine Learning field, which can be briefly defined as enabling computers make successful predictions using past experiences, has exhibited an impressive development recently with the help of the rapid increase in the storage capacity and processing power of computers.

A book recommendation system is a type of recommendation system where we have to recommend similar books to the reader based on their interest. The book recommendation system is used by online websites which provide e-books like Google play books, open library, good Read's, etc. A recommendation system helps an organization to create loyal customers and build trust by providing them personalized products and services for which they to the site.

At the end of the training phase of the internship, this project 'Book Recommendation System' was taken up. The main objective was to apply the theoretical concepts and successfully build a model to recommend similar books.

# TABLE OF CONTENTS

# LIST OF FIGURES

# CHAPTER 1
# INTRODUCTION

A recommendation engine is a class of machine learning which offers relevant suggestions to the customer. The recommendation system today are so powerful that they can handle the new customers too who has visited the site for the first time. A recommendation system is usually built using three techniques which are content-based filtering, collaborative filtering and a combination of both.

This project uses Content based filtering methods based on the description of an item and the user's preferred choices. K-Nearest Neighbors is incorporated to find similar books in terms of Language, Author and rating and build the recommender system. It is one of the simplest Machine Learning algorithms based on Supervised Learning technique. This algorithm assumes the similarity between the new case/data and the available cases and put the new case into the category that is most similar to the available categories. It stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category. It does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. At the training phase, it just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

Moreover, this project uses certain libraries for data exploration, in order to get highly rated books, best-selling Authors and many more. It plots the bar graphs to get better insights about the features of the dataset. Finally, it trains the model and user's choice of a particular book is passed. It returns 10 other similar books and hence recommender system is built.

# CHAPTER 2
# COMPANY PROFILE

Prinston Smart Engineers was started in 2004 in New Delhi. With Dozens of successful projects under its belt for more than 13 years, it is one of the most trusted Engineering, Maintenance & Training Services in Delhi and also extends its services across India in various states such as Karnataka (Bangalore, Mysore), Gujarat and Jaipur.

The aim to create a responsive client relationship that allows us to meet and even exceed the goals of each project.

In 2016 the company expanded its services in Skill development and Training program for engineering students in various engineering domains. India today produce 1.5 million Engineering graduates a year, it is however agreed by all that 75% of these graduates are unemployed, agreeing to the fact that there is need to provide Skill training to the Engineering students. In 2018 AICTE announced that an Internship is mandatory for engineering students to ensure that technical students get exposure to the Industrial environment, current technology relevant to their subject and opportunities to learn understand and sharpen real time technical and managerial skills. On request from colleges, the company began providing quality internship under the guidance of their expert's team to the students from various universities and colleges. In 2020 Prinston Smart Engineers collaborated with Wedir-Tech Trading Contracting & Services W.L.L, Doha, Qatar with a MoU for their expertise in technical know-how and skill development.

List of function performed-

- Operation & Maintenance: This involves Air Conditioning Equipment Maintenance, Refurbishment and renovation of existing system as well as Preventive maintenance.
- Mechanical and electrical Construction: The company provides services in both large and small projects pertaining to Mechanical and Electrical Construction. This includes Electrical Wiring, Plumbing, Fabrication, Welding and sprinkler systems.
- Skill Development and training: The Company offers various courses and programs in reskilling and training engineering graduates. They have tailor made

programs which are specifically designed to minimize the gap between the industrial demands and young engineers.

- Interior designing: The Company consists of professional interior designers that are qualified by education, experience and examination to enhance the function and quality of interior spaces for the purpose of improving the quality of life, increasing productivity and protecting the health safety and welfare of the public.

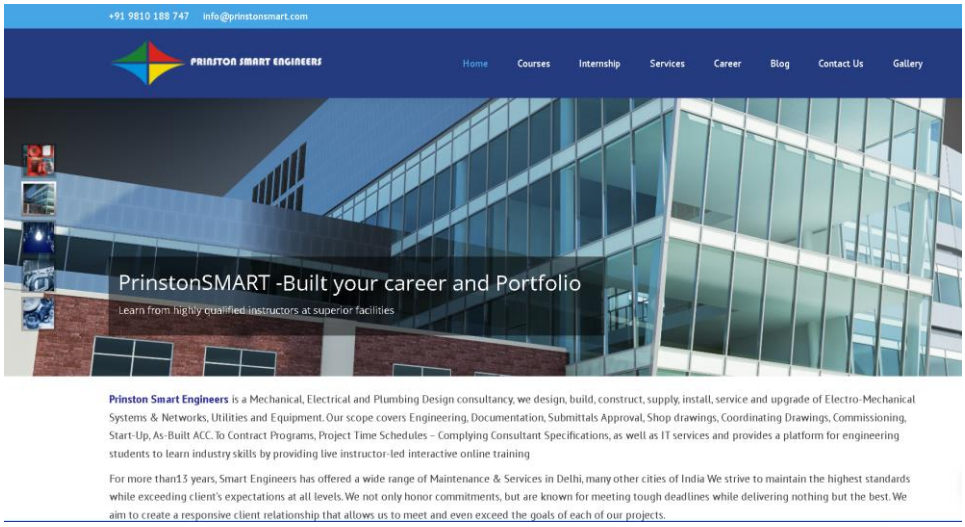


Fig 2.1: Company Website

# CHAPTER 3
# ABOUT THE COMPANY

COMPANY NAME: PRINSTON SMART ENGINEERS

CEO: MR. ASIF AKHTER

CO- FOUNDER: MRS. FARHEEN FARHATH

Prinston Smart Engineers is wholly owned company by Mr. Asif Akhtar. He is sole investor and there are no shareholders or shares since the company is neither listed on National Stock Exchange of India, nor on Bombay Stock Exchange.

Prinston smart engineers is involved in a wide range of service sectors mentioned below:

- Operation & Maintenance.
- Mechanical and Electrical Construction.
- Skill Development and training.
- Interior Designing
- Energy Saving Solutions.
- MEP design & consultancy services.

With respect to their Operations, Maintenance and MEP division, customers of the company include clients such as SHIRPA Group, Honeywell, C&S electricals and ITC Hotels. Along with these, the company has worked with several other organisations in providing interior designing and energy saving solutions. From their Skill Development and training division's perspective, the customers belong to a wide variety of engineering students from over 50+ colleges based in and around Bangalore as well training corporate professionals as well.

Prinston Smart Engineers has 4 major divisions in the organisation:

- Construction department: has 15 employees.
- Maintenance department: has 15 employees.
- Designing department: has 15 employees.
- Training department: has 20 employees.
- Altogether, Prinston Smart Engineers has employed over 65 employees of which 20 employees are situated in India and 45 in are based in Qatar.

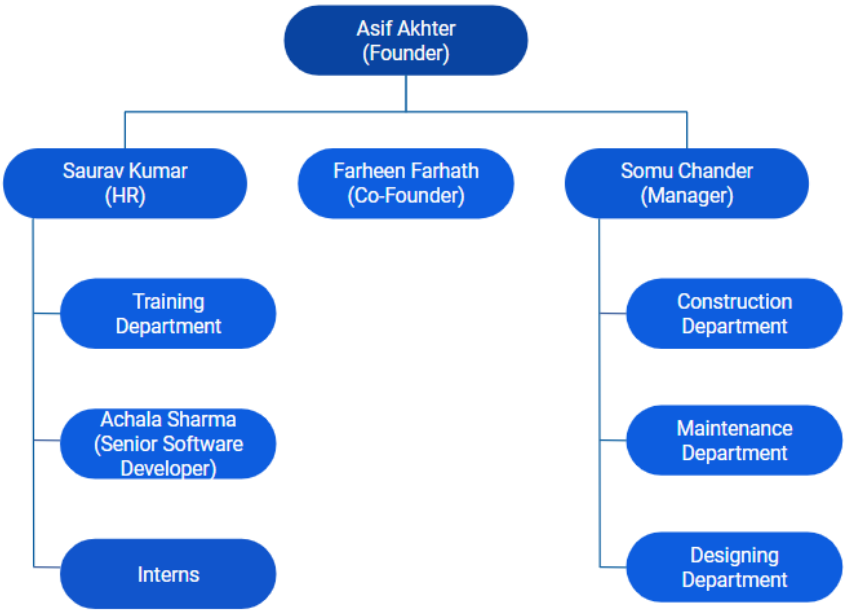


Fig 3.1: Organization Flow Chart

# CHAPTER 4

# TASKS PERFORMED

The objective of this internship was to build 'Book Recommendation System' in the project phase. The tasks performed in the project are described below:

## 4.1 Data acquisition

- Data acquisition Machine learning needs two things to work, data and models.
- When acquiring the data, one must be sure to have enough features (aspect of data that can help for classification, like the book genre) populated to train correctly your learning model.
- The dataset was downloaded from Kaggle website.
- It contains 11,126 rows and 12 columns.
- The features of the dataset includes : BookID, Title, Authors, Average Rating, Isbn code, Language code, Number of pages, Total Ratings, Total text reviews, Publication date, Publisher.

## 4.2 Data Pre-Processing

- Data preparation involves five sub-processes to be followed.
- They are selection, cleansing, construction, integration, and formatting of data.
- In other words, all these steps comprise all the activities that must be performed for construction of the final data set.
- The real-world data often has a lot of missing values. The cause of missing values can be data corruption or failure to record data. The handling of missing data is very important during the preprocessing of the dataset as many machine learning algorithms do not support missing values.
- When rating counts column had missing values, mean was used to calculate rating approximately.
- Label Encoding refers to converting the labels into a numeric form so as to convert them into the machine-readable form. Machine learning algorithms can then decide in a better way how those labels must be operated.
- Language code is encoded using label encoder.

- Normalization and scaling is also performed to obtain the values within the same range. Normalization adjusts the values of your numeric data to a common scale without changing the range whereas scaling shrinks or stretches the data to fit within a specific range.

## 4.3 Data Analysis

- Exploratory data analysis is an approach to analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods.
- The main purpose of EDA is to help look at data before making any assumptions. It is used to understand data, get some context regarding it, understand the variables and the relationships between them, and formulate hypotheses that could be useful when building predictive models.
- Different bar graphs are plotted to interpret the data.
- A graph for – Top 10 authors with most books is plotted where authors are marked on y-axis and number of books are plotted on x-axis.
- Another graph to view the books with maximum rating is plotted where average rating is marked on y-axis and book ID on x-axis.

## 4.4 Model

- K-nearest model is used to find similar books. It finds closest books and recommend to the user. The algorithm is as follows:
    - Load the data.
    - Initialize K to the chosen number of neighbours.
    - For each example in the data, calculate the distance between the query example and the current example from the data.
    - Add the distance and the index of the example to an ordered collection.
    - Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances.
    - Pick the first K entries from the sorted collection.
    - Get the labels of the selected K entries and return the mode of the K labels.

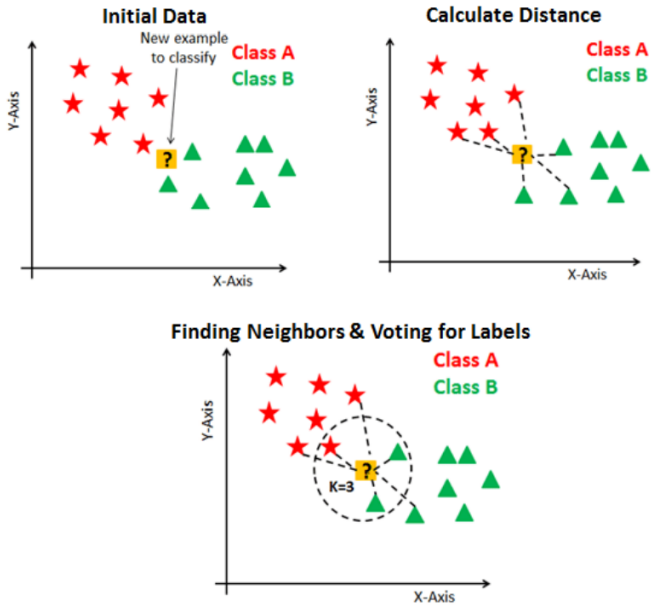


Fig 4.4: Working of KNN algorithm

# CHAPTER 5

# REFLECTION

During the internship period, I had hands-on experience with many real-time projects. By working with numerous datasets, I got deeper understanding on how to understand, analyze and handle data. This internship has provided me a great experience, lessons and the tools that I will need to get a gig in the future. Moreover, it has given me the opportunity to grow and learn fully on how these concepts can be used to solve real-time problem statements. Some of the other skills gained are listed below:

- Responsibility and keeping commitments: I had to stay focused and committed to attend all the training sessions. In addition, there was a deadline within which the project along with report and presentation had to be submitted. By this, I had to work on the project sincerely with the time frame in mind.
- Verbal and written Communication: Communication in a formal way while interacting with the trainer. Writing skills improved through making the report and presentation.
- Problem solving: Introduced to real-life work problems and worked on regular tasks. Problem solving skill has improved by working on the challenging project as well. Handling and solving all errors was experienced.
- Work Ethics: Attending training sessions regularly and completing the projects and other tasks within the stipulated time without making excuses. Punctuality and Time Management was also enhanced

# CHAPTER 6

# RESULTS

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11127 entries, 0 to 11126
Data columns (total 12 columns):
 #   Column              Non-Null Count  Dtype 
---  ------              --------------  ----- 
 0   bookID              11127 non-null  int64 
 1   title               11127 non-null  object
 2   authors             11127 non-null  object
 3   average_rating      11127 non-null  object
 4   isbn                11127 non-null  object
 5   isbn13              11127 non-null  object
 6   language_code       11127 non-null  object
 7   num_pages           11127 non-null  object
 8   ratings_count       11127 non-null  int64 
 9   text_reviews_count  11127 non-null  int64 
 10  publication_date    11127 non-null  object
 11  publisher           11127 non-null  object
dtypes: int64(3), object(9)
memory usage: 1.0+ MB
```

**Fig 6.1: Overview of Dataset**

```
plt.figure(figsize=(8,3))
ax=sns.barplot(book_aut['title'], book_aut.index, palette='inferno')
ax.set_title("Top 10 authors  with most books")
ax.set_xlabel("Total number of books")

total=[]

for i in ax.patches:
    total.append(i.get_width())

totals=sum(total)

for i in ax.patches:
    ax.text(i.get_width()+.2, i.get_y()+.2, str(round(i.get_width())))

plt.show()
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables as keyword args: x, y.
  FutureWarning
```

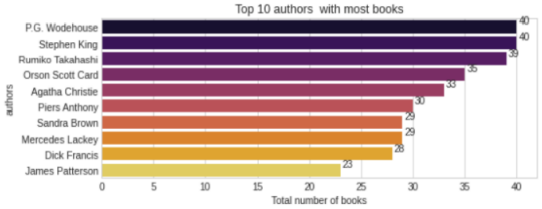


**Fig 6.2: Data exploration: Top 10 Authors**

```
[63] plt.figure(figsize=(8,3))

     ax=sns.barplot(most_rated['ratings_count'], most_rated.index, palette='inferno')
     ax.set_title("Top 10 books  that are most rated")


     total=[]

     for i in ax.patches:
         total.append(i.get_width())

     totals=sum(total)

     for i in ax.patches:
         ax.text(i.get_width()+.2, i.get_y()+.2, str(round(i.get_width())))

     plt.show()
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables as keyword args: x, y.
  FutureWarning
```

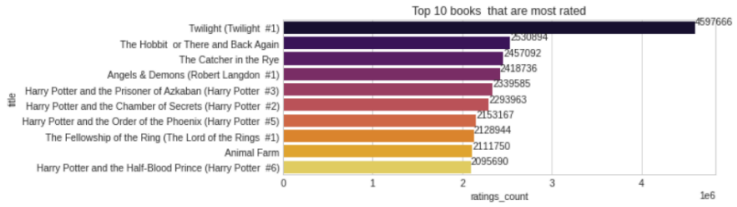


**Fig 6.3: Data exploration: Top 10 Books**

```
[38] df2=df.copy()
```

```
df2.loc[(df2['average_rating']>=0)&(df2['average_rating']<=1),'rating between']="between 0 and 1"
df2.loc[(df2['average_rating']>1)&(df2['average_rating']<=2),'rating between']="between 1 and 2"
df2.loc[(df2['average_rating']>2)&(df2['average_rating']<=3),'rating between']="between 2 and 3"
df2.loc[(df2['average_rating']>3)&(df2['average_rating']<=4),'rating between']="between 3 and 4"
df2.loc[(df2['average_rating']>4)&(df2['average_rating']<=5),'rating between']="between 4 and 5"
```

```
[40] df2.sort_values('average_rating')
```

```
rating_df=pd.get_dummies(df2['rating between'])
language_df=pd.get_dummies(df2['language_code'])
```

**Fig 6.4: Data Preparation**

```
[47] from sklearn import neighbors
     from sklearn.model_selection import train_test_split

     model=neighbors.NearestNeighbors(n_neighbors=10, algorithm='ball_tree')

     model.fit(feature)

     dist, idlist=model.kneighbors(feature)
```

**Fig 6.5: Training KNN model**

```
Book_name=bookRecom('Harry Potter and the Half-Blood Prince (Harry Potter  #6)')
Book_name
```

```
['Harry Potter and the Half-Blood Prince (Harry Potter  #6)',
 'The Feelings Book: The Care & Keeping of Your Emotions',
 'The Complete Shorter Fiction of Virginia Woolf',
 'Warrior of the Light',
 'The Known World',
 'Hour Game (Sean King & Michelle Maxwell  #2)',
 "The Art of Innovation: Lessons in Creativity from IDEO  America's Leading Design Firm",
 'Blink: The Power of Thinking Without Thinking',
 'Candide',
 'Beach Blanket Bad Boys']
```

**Fig 6.6: Recommendation 1**

```
[52] Book_name=bookRecom('Democracy in America')
     Book_name
```

```
['Democracy in America',
 'Lost Illusions (La Comédie Humaine)',
 'Underworld Unleashed',
 'Katherine',
 'La reina de los condenados (Crónicas Vampíricas  #3)',
 'Julius Caesar',
 'A Secret Splendor',
 'Cat Laughing Last (Joe Grey  #7)',
 'Art Objects: Essays on Ecstasy and Effrontery',
 'Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought']
```

**Fig 6.7: Recommendation 2**

# CONCLUSION

Machine learning is quickly growing field in computer science. It has applications in nearly every other field of study and is already being implemented commercially because machine learning can solve problems too difficult or time consuming for humans to solve.

As a result, we have studied the future of Machine Learning. Also, study algorithms of machine learning. Along with we have studied its application which will help you to deal with real life. Common methods and popular approaches used in the field, suitable machine learning programming languages, and also covered some things to keep in mind in terms of unconscious biases being replicated in algorithms.

This has set a direction and a route which will definitely boost my career. I worked on this challenging project independently. Moreover, it has made me take quite a few decisions alone such as deciding the dataset and the approach that this project requires.

Having successfully built this recommendation system, which could be used in several sites, I feel a sense of accomplishment

# REFERENCES

[1] https://jupyter.org/

[2] https://www.kaggle.com

[3] https://towardsdatascience.com

[4] https://stackoverflow.com

[5] www.geeksforgeeks.org

[6] www.javapoint.com

[7] www.coursera.com