Course Recommendation System

Report submitted in partial fulfillment of the requirement for the degree of

B.Tech

in

Computer Science & Engineering

ANd9GcRcmIsloy3g35B_JsRNO8FT1UVoHApPujW-ZfaOSnKNCyg35wmufV3iEg

Under the supervision

of

Ms. Charu Gupta

Assistant Professor

Department of Computer Science & Engineering

by

Achintya Sarkar 00720802715 , CSE-A

Himanshu Ratnani 03220802715 , CSE-A

Aayush Tiwari 00420802715 , CSE-A

Abhishek Jain 0062002715, CSE-A

Department of Computer Science & Engineering

Bhagwan Parshuram Institute of Technology

PSP-4, Sec-17, Rohini , Delhi-89

April 2019

**DECLARATION**

This is to certify that Report titled “Course Recommendation System”, is submitted by us in partial fulfillment of the requirement for the award of degree B.Tech. in Computer Science & Engineering to BPIT, GGSIP University, Dwarka, Delhi. It comprises of our original work. The due acknowledgement has been made in the report for using others work.

**Date: 22-04-2019** Achintya Sarkar 00720802715 , CSE-A

Himanshu Ratnani 03220802715 , CSE-A

Aayush Tiwari 00420802715 , CSE-A

Abhishek Jain 0062002715, CSE-A

**Certificate by Supervisor**

This is to certify that Report titled “Course Recommendation System” is submitted by Abhishek Jain 00620802715, Achintya Sarkar (00720802715) , Himanshu Ratnani 03220802715, CSE-A , Aayush Tiwari 00420802715 , in partial fulfillment of the requirement for the award of degree B.Tech in Computer Science & Engineering to BPIT, GGSIP University, Dwarka, Delhi. **It is a record of the candidates own work carried out by them under my supervision.** The matter embodied in this Report is original and has not been submitted for the award of any other degree.

(Signature)

**Date 22-04-19 Ms. Charu Gupta**

**Certificate by HOD**

This is to certify that Report titled “Course Recommendation System ” is submitted by Abhishek Jain 00620802715, Achintya Sarkar (00720802715), Himanshu Ratnani (03220802715), CSE-A , Aayush Tiwari 00420802715, under the guidance of Ms. Charu Gupta in partial fulfillment of the requirement for the award of degree B.Tech in Computer Science & Engineering to BPIT, GGSIP University, Dwarka, Delhi. The matter embodied in this Report is original and has been dully approved for the submission.

  (Signature)

**Date:22-04-19 Dr. Deepali Virmani**

**ACKNOWLEDGEMENT**

We would like to express a word of thanks to all those who have provided us with sincere advices and information during the course of the period we thought and worked for the creation of the product. It was indeed a great pleasure for us to work in a very cooperative, enthusiastic and learning atmosphere it was a very informative experience and we all have grown a lot during the course of creation of the final product and are much prepared to take on future projects and complete them before deadlines with utmost sincerity and delinquency.

We would like to express our sincere gratitude to our mentor, Ms. Charu Gupta, assistant professor, C.S.E. Department for guiding us through this project. We thank her for the invaluable guidance and encouragement she provided throughout this project. The valuable suggestions given by her is one of the main reasons for the completion of this project. We wish to express our sincere gratitude to Dr. Deepali Virmani, HOD, C.S.E. Department for her strict supervision, constant encouragement, inspiration and guidance, which ensured the worthiness of our report. We would also like to thank our principal Dr. Payal Pahwa for her support and guidance.

We would also like to show our sincere gratitude to the Management Department, BPIT for giving us such an imperative opportunity to showcase our skills and letting us come up with a solution to a real life problem and having faith in us throughout this project.

**Table of Contents**

**List of Figures ……………………………………...……………………………....…….. ix**

**Abstract ……………………………………………...………………………...…………. x**

**1. Introduction ......................................................................................................................** 1

1.1 Recommendation System ..................................................................................... 2

1.2 Types of recommendation system......................................................................... 3

1.2.1 Collaborative Filtering ………………………………………....…….. 3

1.2.2 Content Based Filtering …………………………………………….… 4

1.2.3 Hybrid Recommendation system ……………………………..…...…. 5

1.3 Need for Project …………………………………………………….………….. 6

1.4 Problem Statement ……………………………………………………..…...….. 6

**2. Related Work ....................................................................................................................** 7

2.1 Machine Learning ................................................................................................. 7

2.2 Natural Language Processing ............................................................................... 8

2.3 K-Means Algorithm ............................................................................................. 11

2.4 Linear Regression ………………………………………………..….….…..….. 12

2.5 Pearson Correlation ……………………………………………………………. 13

2.6 Survey ……………………………………………………………………..…… 13

2.7 Work Done ……………………………………………………………………... 15

**3. System analysis and Design .............................................................................................** 17

3.1 Software Requirement Specification (SRS) ......................................................... 17

3.1.1 Definition …………………………………………………………….. 17

3.1.2 Functional Requirements …………………………………………….. 18

3.1.3 Hardware Requirements ……………………………………………... 18

3.1.4 Software Requirements ………………………………………………. 19

3.2 Dataset Description …………………………………………………………….. 19

3.2.1 Dataset Extraction ……………………………………………………. 19

3.2.2 Validation of Dataset ……………………………………………….... 19

3.2.3 Data Split Ratio ………………………………………………………. 20

3.3 Threats to Validity …………………………………………………………..…. 21

3.3.1 Mortality …………………………………………………………..…. 21

3.3.2 Maturation ………………………………………………….……….... 21

3.3.3 System Fails ……………………………………………….………..... 21

3.4 Use case Diagram ………………………….……………………...……...……. 21

3.4.1 Use Cases ……………………………………………………..……… 22

3.4.2 Actors ……………………………………………………...…………. 22

3.5 DFD ………………………………………...……………………………..…… 23

3.5.1 DFD Level 0 …………………………………………...…………….. 23

3.5.2 DFD Level 1 ………………………………………………...……….. 23

**4. Proposed Work ................................................................................................................** 25

4.1 Implementation .................................................................................................... 25

4.2 Flow Charts .......................................................................................................... 26

4.2.1 Course of Action ………………………………………………...…… 28

4.2.2 Algorithm …………………………………………………………….. 30

**5. Implementation and Results ........................................................................................... 31**

5.1 Working Explanation ……………………………………………………..……. 31

5.2 Preprocessing of data ……………………………………………………….….. 32

5.3 Making the feature set ……………………………………………….…………. 34

5.4 Training process …………………………………………………….………..… 37

5.5 User item matrix ……………………………………………………….………..39

5.6 Creating a scoring mechanism ……………………………………….……...…. 41

5.7 Results ………………………………………………………………….………. 42

5.8 Web Application …………………………………………………………….…. 44

**6. Conclusion .......................................................................................................................** 45

**7. Future Work .....................................................................................................................** 47

**References ………………………………………………………......…….….……...……..** 48

**Appendix …………………………………………………………….….…….…….………**49

**List of Figures**

Figure 1.1: Collaborative based filtering system .................................................................... 3

Figure 1.2: Content based filtering system ............................................................................. 4

Figure 1.3: Hybrid based recommendation system ................................................................. 5

Figure 2.1: Clustering after Iteration #1 ………………………………………………..….. 12

Figure 3.1: Interaction of recommendation system ............................................................... 22

Figure 3.2: DFD level 0 …………………………………………………….……..……….. 23

Figure 3.3: DFD level 1 ………………………………………………………….…..….…. 24

Figure 4.1: Training dataset flow chart ……………………………………………………. 27

Figure 4.2: Course of action for collaborative filtering ……………………………………. 29

Figure 5.1: Set of reviews loaded in ipython notebook ……………...………….…….…… 31

Figure 5.2: Function which makes only English reviews available ……………….….…… 32

Figure 5.3 Enchant Function ……………………………………………………………….. 33

Figure 5.4 Word of Reference ……………………………………………………….…..… 33

Figure 5.5 Indexes to drop non English reviews …………………………………….....….. 33

Figure 5.6 New array of English Reviews …………………………………………………. 34

Figure 5.7: Process of making the features …………………………………………...…… 34

Figure 5.8: Get all the review of considered courses ……………………………….…..…. 35

Figure 5.9: Tokenize each sentence …………………………………………………....….. 35

Figure 5.10: Relating words to labels ……………………………………………...………. 35

Figure 5.11: Getting individual words ……………………………………………..…….... 36

Figure 5.12: Removing Stop words ………………………………………………..……… 36

Figure 5.13: Getting the frequency of clean words ……………………………..……….... 36

Figure 5.14: Selecting 3000 words …………………………………………….………….. 37

Figure 5.15: Function to create a 3000 2D list ………………………………...…….……. 37

Figure 5.16: Informative features found during training ……………………..…….…...… 38

Figure 5.17: NLTK inbuilt classifier ………………………………………….……….….. 38

Figure 5.18: Most informative words …………………………………………………….. 38

Figure 5.19: Classifier function …………………………………………………….…….. 39

Figure 5.20: Store numberical data ….……….…….….……………………….………… 39

Figure 5.21: User item matrix for collaborative filtering ……………………….….……. 40

Figure 5.22: Calculating scores ……………………………………………………...…… 40

Figure 5.23: Set of questions asked from the new user ………………………..………… 41

Figure 5.24: Results of recommendation system ………………………………….….….. 42

Figure 5.25: Data Slicing …………………………………………………………...……. 42

Figure 5.26: Function to get details about the user ………….…………………………… 43

Figure 5.27: Get recommendation function …………………………………………..….. 43

Figure 5.28: Course Recommendation Results ………………………….……………….. 43

Figure 5.29 Welcome page of web application …………………………….……….…… 44

Figure 5.30 Questionnaire ………………………………………………………...……… 45

Figure 6.1 Comparison of Distance Metrics for the Proposed Approach ………………….46

**Abstract**

Choosing a course from a large list of courses can prove to be a challenging job for a student. In order to solve this problem the proposed approach uses a recommendation system which plays a significant role in suggesting the best suitable course to the student according to his/ her personal learning ability. The ideology is that the system recommends the courses based on the similarities of the students who want the recommendation with the students who have already taken and rated the courses earlier, rather than basing the recommendations purely on a knowledge base. In the case of new students prior to suggesting a course the approach uses some questionnaire and a set of parameters for developing a knowledge base of these students so that the recommendation can be optimized. The initial approach used is collaborative based, where the user-item matrix is considered for getting the score of the user and his similar users. The proposed recommender system performs better by mitigating the weakness of basic individual recommender systems along with improving the classifiers used in real-time. This is primarily done by analyzing each user currently using the system. **Keywords: Course recommendation system, Collaborative, User-Item Matrix.**