**CLOUD COUNSELAGE**

**PVT. LTD.**



PROJECT REPORT

**EVENT RECOMMENDER SYSTEM**

By

Chaitanya Deshpande

# ACKNOWLEDGEMENT

It gives me pleasure in presenting the project report on ‘Event Recommender System’.

Firstly, I would like to express my indebtedness appreciation to my guide Aarya Bodhankar. His constant guidance and advice played very important role in successful completion of the report. He always gave me his suggestions, that were crucial in making this report as flawless as possible.

I would like to express our gratitude towards Mr. Jayanth G S for his kind co-operation and encouragement which helped me during the completion of this report.

Also, I wish to thank Mr. Tushar Topale, Mr. Nirbhey Singh Pahwa and all members of Cloud Counselage Pvt. Ltd. for their whole hearted co-operation during this internship.

Last but not the least, the backbone of my success and confidence lies solely on blessings of dear parents and lovely friends.

# TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **CONTENTS** | **Page No.** |
| **1** | **Introduction** | **3** |
| **2** | **Problem Statement** | **4** |
| **3** | **Project Overview** | **5** |
| **4** | **Recommender System** | **6** |
| **5** | **Multiclass Text Classification** | **6** |
| **6** | **Libraries Used** | **8** |
| **7** | **Functional Process flow** | **9** |
| **8** | **Conclusion** | **10** |

# INTRODUCTION

Recommender System is a system that seeks to predict or filter preferences according to the user’s choices. Recommender systems are utilized in a variety of areas including movies, music, news, books, research articles, search queries, social tags, and products in general.

Recommender systems produce a list of recommendations in any of the two ways –

* **Collaborative filtering:**Collaborative filtering approaches build a model from user’s past behaviour (i.e. items purchased or searched by the user) as well as similar decisions made by other users. This model is then used to predict items (or ratings for items) that user may have an interest in.
* **Content-based filtering:**Content-based filtering approaches uses a series of discrete characteristics of an item in order to recommend additional items with similar properties. Content-based filtering methods are totally based on a description of the item and a profile of the user’s preferences. It recommends items based on user’s past preferences.

# PROBLEM STATEMENT

The aim is to create a Recommender System that recommends only relevant events to each

employee and intern based on their preferences, whenever the company receives invites for said

events. The system should read new events and autonomously classify them into various

domains. It should then match the event with all of those in the company database who have

given said domains as a preference. Finally, for each event, the system should output the list of

people whose preferences match with the event’s detected domain. Create a report

documenting your approach and methodology followed.

**OBJECTIVES –**

• To read a set of events as input.

• To classify each event into one or more domains.

• To fetch the employee database with domain and event preferences.

• To match each event with all interested employees.

• To output the list of matching employees per event.

# PROJECT OVERVIEW

This Project is to develop recommendation system for Internal employees of company. System will read various events that has been received in the file format. System will classify the input events based on domain and type of events (Webinar, Seminar, certificate course, etc.). System will read data from list of employees and with help of its own intelligence it will recommend events to the employees. Output file will be generated by system which will have event data and event wise interested employees. This recommender system will be developed by using python.

# Recommender Systems

Recommender Systems (RSs) are software tools and techniques providing suggestions for items to be of use to a user or a group of users. The RSs provides support for the process of decision-making for a user or a group of users. Recommendations are predictions of the most suitable items based on user's preferences. An Item is the general term referring to what the system recommends to users.RS normally focuses on a specific type of item (e.g., product, news or command) and that the user interface has to be designed accordingly.

**Types of RSs**

Recommender systems can implement several types of filtering, following the sources of information and algorithm used for the filtering. These filtering can be:

* collaborative filtering,
* content-based filtering
* knowledge-based filtering
* hybrid filtering

Content-based filtering approaches uses a series of discrete characteristics of an item in order to recommend additional items with similar properties. Content-based filtering methods are totally based on a description of the item and a profile of the user’s preferences. It recommends items based on user’s past preferences.

Considering various solutions for recommender system like Content based filtering and collaborative filtering for given problem content-based filtering was best choice. Because, in given employee data the domain and event choice of employees are specified. For this scenario the Content based filtering is best suitable choice.

# Multiclass text classification

System will take event description or event info as an input. The employees in given dataset has various domain and event types as preferences hence the input must be classified in domain and event types. For the multiclass classification various classifiers like Naive Bayes, Linear Regression, Random Forest are available nut based on their accuracy score Linear SVM Classifier gives the best classification for multiclass classification. So, based on this approach of content-based filtering and SVM as a classifier system will ensure the expected accuracy of recommendations.

In SVM we can classify linear as well as non-linear data. The nonlinear data will be represented into n-dimensional space using kernel tricks. Then maximum-margin hyperplane will be generated to classify data into classes.

The classifiers and learning algorithms cannot directly process the text documents in their original form, as most of them expect numerical feature vectors with a fixed size rather than the raw text documents with variable length. Therefore, during the preprocessing step, the texts are converted to a more manageable representation.

One common approach for extracting features from text is to use the bag of words model: a model where for each document, a complaint narrative in our case, the presence (and often the frequency) of words is taken into consideration, but the order in which they occur is ignored.

Specifically, for each term in our dataset, we will calculate a measure called Term Frequency, Inverse Document Frequency, abbreviated to tf-idf. To train supervised classifiers, we first transformed the input into a vector of numbers. We explored vector representations such as TF-IDF weighted vectors.

After having this vector representations of the text, we can train supervised classifiers to train unseen data and predict the class in which they classify.

Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them.

Similarity = (A.B) / (||A||.||B||) where, A and B are vectors.

Mathematically, it measures the cosine of the angle between two vectors projected in a multi-dimensional space. In this context, the two vectors I am talking about are arrays containing the word counts of two documents.

The cosine similarity is advantageous because even if the two similar documents are far apart by the Euclidean distance because of the size (like, the word ‘cricket’ appeared 50 times in one document and 10 times in another) they could still have a smaller angle between them. Smaller the angle, higher the similarity.

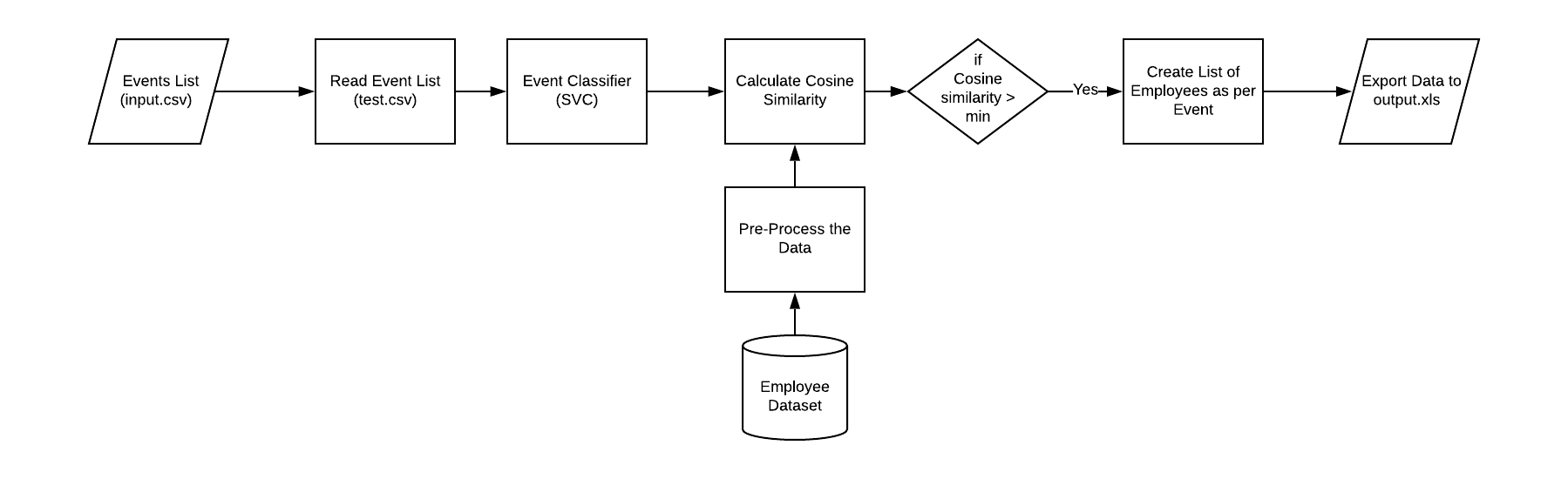
# Libraries Used

• Pandas Library - pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

• Scikit-Learn Library - Scikit-learn is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

• xlsxwriter library - It is library for python that provides facility to export data in MS Excel Format(.xlsx).

# Functional Process Flow



1. System will take number of events as input from csv file and import it into pandas Data Frame.
2. Each event from input file will be classified into various classes according to domain and events using SVM as Classifier. In SVM we can classify linear as well as non-linear data. The nonlinear data will be represented into n-dimensional space using kernel tricks. Then maximum-margin hyperplane will be generated to classify data into classes.
3. The class (Domain and event type) of the input event will be checked with employees’ preferences using cosine similarity. Class of the input event (eg. Machine Learning Certifications) and Employees’ selected domain and event type will be converted into vectors using CountVectorizer. Cosine similarity method calculates cosine angle between those 2 vectors.
4. Based on cosine similarity score of employees, the list employees with maximum score will be stored in list.
5. Then list of employees with maximum score will be stored into MS Excel file using xlsxwriter library.

# Conclusion

Considering various solutions for recommender system like Content based filtering and collaborative filtering for given problem content-based filtering was best choice. Because, in given employee data the domain and event choice of employees are specified. For this scenario the Content based filtering is best suitable choice.