Cloud Counselage Pvt. Ltd.

Internship Program

Live Project

Event Recommendation System for Employees

Author

Vinit Kanse

Acknowledgement

Index

Introduction:

One often misses events of interest sheerly due to a lack of awareness at the right time. Cloud Counselage also receives invites for events in multiple domains that need to be forwarded to people with relevant interests. Hence a recommendation like this helps in companies like Cloud Counselage to recommend events to the employees/interns. 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.

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.

Block diagram-
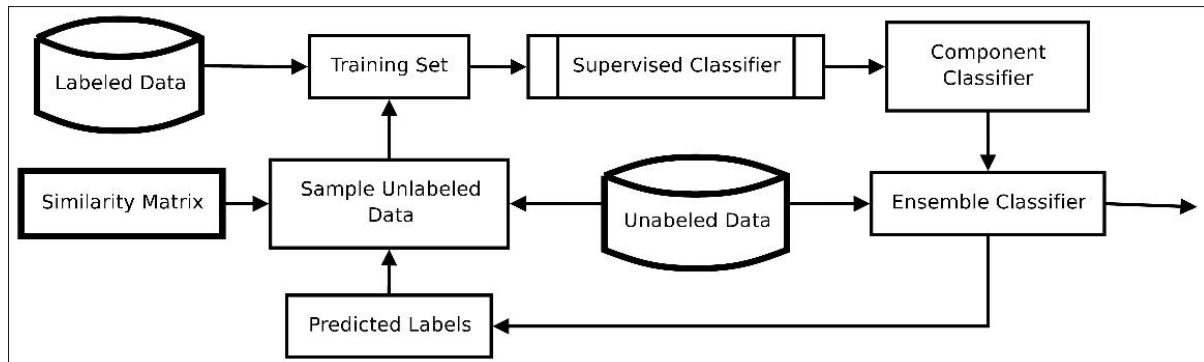
Methodology-

The dataset entitled CCMLEmployeeData.csv is used. This dataset consists of information of the employee names, their domains as well as their events preferences. The input is taken from the user and the domain and events are taken from the training data base. This information is then converted into required format and fed to the system to check for similarity along with the other pre-recorded data of employees. . Using the cosine similarity metric of SciKit Learn library the features between the input sentence, the input events and the event preferences of the employees are determined using the ML approach of Content Based Recommender Systems.

As a result, a list of all the employees for which the events shall be recommended is obtained as output.

Constraints

1) The system is to be created in Python 3.6 or above.
2) The events and employee list to be read shall be provided as a [txt] with 1 element per line.
3) You are allowed to use any ML approach, publicly available resources, and additional training data, but the final output should be on the provided list.
4) The output shall be in the form of a spreadsheet having a column for event name and another for the names of people to recommend said event to.
5) The output spreadsheet should be auto generated by the system you create.

Technology/Requirements

Hardware Requirements: -

RAM- 2GB

Processor- Intel i5

Hard disk- 32 GB minimum

Operating system: Windows, Internet browser

Software used:

Python 3.6

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax

emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

Jupyter Notebook

The **Jupyter Notebook** is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

Python Libraries used-

1) Pandas
   Pandas is a high-level data manipulation tool developed by Wes McKinney. It is built on the NumPy package and its key data structure is called the Data Frame. Data Frames allow you to store and manipulate tabular data in rows of observations and columns of variables.

2) NumPy
   NumPy is a **python** library used for working with arrays. It also has functions for working in domain of linear algebra, Fourier transform, and matrices. **NumPy** was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely.

3) Scikit learn-
   Scikit-learn is a free machine learning library for Python. It features various algorithms like support vector machine,

random forests, and k-neighbours, and it also supports Python numerical and scientific libraries like NumPy and SciPy.

4) Matplotlib-

Matplotlib is one of the most popular Python packages used for data visualization. It is a cross-platform library for making 2D plots from data in arrays. It provides an object-oriented API that helps in embedding plots in applications using Python GUI toolkits such as PyQt, WxPythonotTkinter. It can be used in Python and IPython shells, Jupyter notebook and web application servers also.

5) TF-IDF Vectorizer –

Transforms text to feature vectors that can be used as input to estimator. vocabulary_ Is a dictionary that converts each token (word) to feature index in the matrix, each unique token gets a feature index. ... In each vector the numbers (weights) represent features tf-idf score.

6) CHI2 –

Compute chi-squared stats between each non-negative feature and class. This score can be used to select the n features features with the highest values for the test chi-squared statistic from X, which must contain only non-negative features such as Booleans or frequencies (e.g., term counts in document classification), relative to the classes.

Conclusion

We can build an event recommendation system based on taking an input of string events and classifying them into respective domain and event category using Content Based Filtering and Natural Language Processing. This classification can be further be used to generate the final output csv.

References

1. https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.chi2.html
2. https://developers.google.com/machine-learning/recommendation/content-based/basics
3. https://medium.com/@armandj.olivares/building-nlp-content-based-recommender-systems-b104a709c042