**PROJECT REPORT**

**On**

**Machine Learning Internship**

**By: HRITIK SHARMA**

**Event Recommender System Project For Employees**

**ACKNOWLEDGEMENT**

I take great pleasure in submitting this project report based on “Event Recommendation System for Employees”. I wish to express true sense of gratitude towards HR Harshada Topale and the Administrator of IT Department Mr. Nirbhey Singh Pahwa, for shortlisting me to work upon this internship and Project Manager. I express my heartfelt gratitude towards Mr. Jaynath G S who at very discrete stages in study of this project, contributed their valuable guidance and helped to solve every problem that arose. I also thank the founder of Cloud Counselage Pvt. Ltd. Mr. Tushar Topale, for providing this golden opportunity to work on such an informative internhsip.

I would like to thank the entire team of Cloud Counselage Pvt. Ltd. without whom the completion of this internship would have been impossible.

**TABLE OF CONTENTS**

|  |  |
| --- | --- |
| **Sr. No.** | **CONTENTS** |
| **1** | **Problem Statement** |
| **2** | **Platform/Technology** |
| **3** | **Description and Block Diagram** |
| **4** | **Tools and Libraries Used** |
| **6** | **Recommended Future Improvements** |

# PROBLEM STATEMENT

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

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

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 ofpeople 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.

# PLATFORM/TECHNOLOGY

**Python 3.6:** Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace.

**Anaconda**: Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing, that aims to simplify package management and deployment. The distribution includes data-science packages suitable for Windows, Linux, and macOS.

**Jupyter Notebook**: Project Jupyter is a nonprofit organization created to "develop open-source software, open-standards, and services for interactive computing across dozens of programming languages". Spun-off from IPython in 2014 by Fernando Pérez,

**Hardware and Software Requirements**

Hardware Requirements

* + Operating Systems: Windows/Linux
  + RAM: 4GB and more
  + Hard Disk: 2 GB of HDD
  + Processors: Pentium 3 2.4 GH

Software requirements:

1. Operating System- Windows/Linux
2. Python 3.6
3. IDE- Preferablly Anaconda
4. Internet Browser: Chrome, Mozilla Firefox, Internet Explore
5. any code editor

**DESCRIPTION**

Create a Recommender System that recommends only relevant events to eachemployee and intern based on their preferences, whenever the company receives invites for saidevents. The system should read new events and autonomously classify them into variousdomains. It should then match the event with all of those in the company database who havegiven said domains as a preference. Finally, for each event, the system should output the list ofpeople whose preferences match with the event’s detected domain.

1. **BLOCK DIAGRAM**

Import the Input Text CSV File

Import all the important

Libraries

END

Store the result in xlsx file

Extract the Recommended Employees name

Match the list with the Dataset

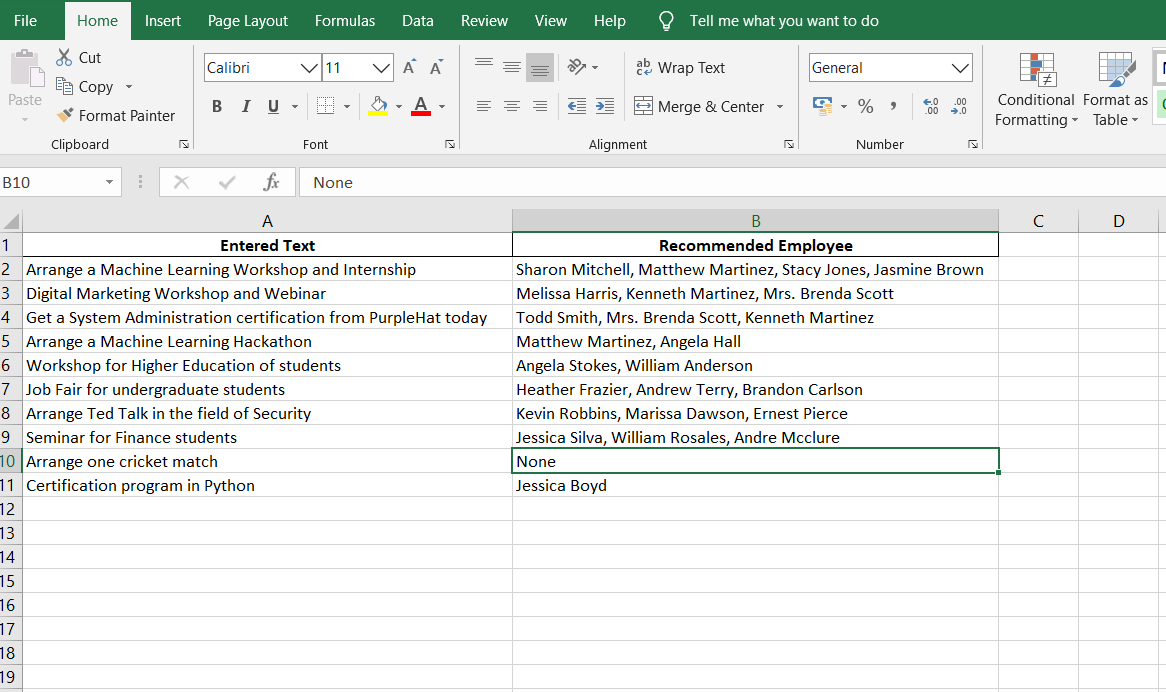
Store the result as Domain and Event list

Text matching of the Dataset using NLP

Import the Dataset

START

1. **RESULT**



1. **ADVANTAGES & DISADVANTAGES**
2. **ADVANTAGES:**

* Easy to arrange events.
* Easy to get good instructor for events.
* Time Saving

1. **DISADVANTAGES:**

* User have to enter specific domains which are mentioned in the Employees CSV File.

1. **APPLICATION**

* Full event organizations and employee related work
* Finding suitable employees.

# TOOLS AND LIBRARIES USED

## Tools:

1. Python:

Python is a high-level, interpreted, interactive and object-oriented scripting language. Due to the complex, scientific computing nature of machine learning applications, Python is considered the most suitable programming language. This is because of its extensive and mature collection of mathematics and statistics libraries, extensibility, ease of use and wide adoption within the scientific community. Python is designed to be highly readable.

As a result, Python has become the recommended programming language for machine learning systems development. A Data Scientist possesses knowledge of solutions to various classes of data-oriented problems and expertise in applying the necessary algorithms, statistics, and mathematical models, to create the required solutions. Python is recognized among the most effective and popular tools for solving data science related problems.

1. Microsoft Excel:

A CSV is a comma-separated values file, which allows data to be saved in a tabular format. CSVs look like a garden-variety spreadsheet but with a csv extension. CSV files can be used with most any spreadsheet program, such as Microsoft Excel or Google Spreadsheets. Datasets are usually of the csv format.

## Libraries:

1. Pandas:

* Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. The name Pandas is derived from the word Panel Data – an Econometrics from Multidimensional data.
* In 2008, developer Wes McKinney started developing pandas when in need of high performance, flexible tool for analysis of data.
* Prior to Pandas, Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data — load, prepare, manipulate, model, and analyze.
* Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

1. Numpy:

* NumPy is a very popular python library for large multi-dimensional array and matrix processing, with the help of a large collection of high-level mathematical functions.
* It is very useful for fundamental scientific computations in Machine Learning.
* It is particularly useful for linear algebra, Fourier transform, and random number capabilities.
* High-end libraries like TensorFlow uses NumPy internally for manipulation of Tensors.

1. Sklearn:

* Scikit-learn is one of the most popular ML libraries for classical ML algorithms.
* It is built on top of two basic Python libraries, viz., NumPy and SciPy.
* Scikit-learn supports most of the supervised and unsupervised learning algorithms.
* Scikit-learn can also be used for data-mining and data-analysis, which makes it a great tool who is starting out with ML.

# RECOMMENDED FUTURE IMPROVEMENTS

1. Currently System is on command line interface and required packages need to be download to fully run the system. So GUI interface can be made for the system for great user experience and mail can be sent to recommended employees
2. More data can be used to improve accuracy