INTERNSHIP

**Learning Objectives/Internship Objectives**

* Internships are generally thought of to be reserved for college students looking to gain experience in a particular field. However, a wide array of people can benefit from Training Internships in order to receive real world experience and develop their skills.
* An objective for this position should emphasize the skills you already possess in the area and your interest in learning more  Internships are utilized in a number of different career fields, including architecture, engineering, healthcare, economics, advertising and many more.
* Some internship is used to allow individuals to perform scientific research while others are specifically designed to allow people to gain first-hand experience working.
* Utilizing internships is a great way to build your resume and develop skills that can be emphasized in your resume for future jobs. When you are applying for a Training Internship, make sure to highlight any special skills or talents that can make you stand apart from the rest of the applicants so that you have an improved chance of landing the position.

**Chapters :1**

Learning Phases

Stage : 1

* Python
* Flask
* GitHub
* Login –Registration System

Stage : 2

* Authentication
* Java Script Basic
* JWT
* JQuery
* Create Simple Website

Stage : 3

* Pandas
* Numpy
* Matplotlib
* BS4
* Selenium

Stage : 4

* Advance Pandas
* Sk-learn
* NLTK

Stage : 5

* Manipulating All
* Frontend
* Managing Routs for Frontend in Flask
* Debug
* Customizing

Stage : 6

* Class / Function
* Merge All Small Functions in One Class
  + Sk-Learn
  + NLTK
  + Data Pre-Processing
  + DataFrame Processing
* Create Full Stack Website
* Customizing

Stage : 7

* Web Server
* Gunicorn
* Docker
* VM
* VPS

**Chapters :2**

2.1 Introduction

* Movie recommendation systems are software applications designed to suggest movies to users based on their viewing history, preferences, and behavior. The primary goal of these systems is to personalize movie recommendations for users and improve their overall viewing experience. Movie recommendation systems are widely used by streaming platforms such as Netflix, Hulu, and Amazon Prime Video to provide users with a customized list of movies to watch.
* Another reason for the importance of movie recommendation systems is their ability to increase user engagement and retention. By suggesting personalized recommendations, these systems encourage users to spend more time on the platform and watch more movies.
* Movie recommendation systems use a variety of techniques to suggest movies to users. One of the most popular techniques is collaborative filtering, which involves analyzing the viewing behavior of users to identify patterns and similarities. This technique uses two approaches: user-based and item-based. In the user-based approach, the system identifies users with similar viewing behavior and recommends movies based on their collective behavior. In the item-based approach, the system identifies movies with similar characteristics and recommends movies based on their similarities.
* Overall, movie recommendation systems use a combination of techniques to analyze user behavior and movie features to suggest personalized recommendations to users. By doing so, these systems improve the overall viewing experience of users and increase user engagement and retention on streaming platforms.

2.2 Modules Description

Data Scraping :

Data scraping, also known as web scraping, is the process of extracting data from websites. This involves using automated tools to collect and extract data from web pages, and then saving it in a structured format, such as a spreadsheet or a database.

Pre-Processing of Data:

Pre-processing of data refers to the various steps and techniques used to prepare data for analysis or modeling. Pre-processing is a crucial step in the data science pipeline, as it helps to ensure that the data is accurate, complete, and in a format that is suitable for analysis.

Processing of Data :

Processing of data refers to the various techniques used to analyze, manipulate and extract insights from data, once it has been pre-processed. Processing of data is a crucial step in the data science pipeline, as it enables researchers, analysts and other professionals to uncover patterns, relationships and trends in data that can inform decision-making.

Interaction with user :

Interaction with users is an important aspect of many software applications, as it enables users to provide input, receive feedback, and interact with the system in a meaningful way.

Backend Technologies :

Backend technologies are the technologies and tools used to develop the server-side of software applications. The backend is responsible for managing data, handling user requests, and executing business logic.

Data Base:

A database is a structured collection of data that is stored and organized in a way that allows efficient retrieval and manipulation of that data. Databases are used to store and manage a wide range of information, from customer records and financial data to scientific research and multimedia content

**Chapters :3**

2.1 Data Scraping

Data scraping, also known as web scraping, is the process of extracting data from websites. It involves automated extraction of data from websites using software tools and techniques.

Data scraping is used to collect information from a large number of websites quickly and efficiently. This information can then be used for various purposes, such as market research, price comparison, lead generation, content creation, and more.

Data scraping can be done in several ways, including using specialized software tools, coding scripts to automate the process, or using browser extensions. The extracted data can be saved in various formats, such as CSV, JSON, or XML.

However, it is important to note that not all websites allow data scraping, and some may have legal restrictions on the use of their data. It is crucial to ensure that data scraping is done ethically and legally, and that appropriate measures are taken to protect privacy and data security.

**Beautiful Soup (bs4):**

Beautiful Soup is a popular Python library used for web scraping and parsing HTML and XML documents. It provides a simple way to navigate and search through the parse tree of an HTML or XML document, and extract the data you need.

Beautiful Soup uses a parser to parse the HTML or XML document and build a parse tree, which can then be searched and navigated using various methods and attributes. It provides a wide range of functionalities, such as searching for specific tags or attributes, extracting text, navigating the parse tree, modifying the document, and more.

Beautiful Soup is a great choice for web scraping projects that involve extracting data from static websites that do not require user interaction.

2.2 Pre-Processing of Data

Pre-processing data is an essential step in data analysis that involves cleaning, transforming, and preparing data for analysis. Pandas is a popular Python library used for data manipulation and analysis

* **Handling missing data:**

Missing data can cause issues during data analysis. Pandas provides several methods for handling missing data, including **dropna()** to remove rows with missing data, **fillna()** to fill missing data with a specified value, or **interpolate()** to fill missing data with interpolated values

* **Removing duplicates:**

Duplicate values can affect data analysis by skewing results. Pandas provides a method to remove duplicate values using the **drop\_duplicates()** function.

* **Renaming columns:**

Column names can be modified to be more informative and easier to work with using the **rename()** function.

* **Changing data types:**

Data types can be changed to the appropriate type using the **astype()** function.

2.3 Processing of Data

* Scikit-learn (sk-learn) is a popular machine learning library in Python that provides a wide range of tools for data processing, including data preprocessing, feature extraction, and feature selection. NLTK (Natural Language Toolkit) is a Python library used for natural language processing tasks such as tokenization, stemming, and part-of-speech tagging.

**1).Text preprocessing with NLTK:**

Text data needs to be preprocessed before it can be used in machine learning models. NLTK provides tools for text preprocessing, including tokenization, stemming, and stop-word removal.

**2). Scaling data with sk-learn:**

Machine learning models can be sensitive to the scale of the input features. Sk-learn provides a variety of methods for scaling data, including standardization and normalization.

**3).Encoding categorical variables with sk-learn:**

Categorical variables need to be encoded before they can be used in machine learning models. Sk-learn provides methods for encoding categorical variables, including one-hot encoding and label encoding.

**4).Feature extraction with sk-learn:**

Feature extraction involves transforming raw data into a set of features that can be used in machine learning models. Sk-learn provides methods for feature extraction, including bag-of-words and TF-ID

2.4 Interaction with user

1. **User Interaction with HTML, CSS, and JavaScript**

* In web development, user interaction can be achieved using HTML for page structure and content, CSS for styling and layout, and JavaScript for dynamic and interactive behavior. Examples of user interaction include form submission, button clicks, hover effects, and animations.

1. **User Interaction with Console/Terminal:**

* In console/terminal applications, user interaction can be achieved using command-line arguments, input prompts, and output messages. Examples of user interaction include command-line options, input validation, and progress indicators.

2.5 Backend Technologies

Flask is a lightweight web framework for building web applications using Python. It is considered a "micro" framework, meaning that it provides only the basic tools needed for web development, but is highly customizable and extensible through a variety of third-party extensions.

Flask is built on top of the Werkzeug WSGI toolkit and the Jinja2 template engine, both of which are written in Python. Flask's modular design allows developers to choose only the components they need, making it a flexible choice for a wide range of web applications.

Flask supports a variety of backend technologies, including:

**Databases**:

Flask can be used with a wide range of databases, including SQL databases like SQLite, MySQL, and PostgreSQL, as well as NoSQL databases like MongoDB and Redis. Flask integrates with popular database libraries such as SQLAlchemy and Flask-SQLAlchemy to provide easy-to-use and flexible database management.

**Authentication and Authorization:**

Flask supports a variety of authentication and authorization methods, including basic authentication, token-based authentication using JWT, and OAuth for third-party authentication.

**RESTful APIs:**

Flask can be used to build RESTful APIs, which provide a standardized interface for accessing and manipulating data over HTTP. Flask's lightweight design and flexible routing system make it a popular choice for building APIs.

**Testing:**

Flask provides built-in support for testing, including a built-in test client that allows developers to simulate HTTP requests and test their application's responses. Flask also supports popular testing frameworks like pytest and unittest.

**Caching**:

Flask integrates with caching libraries like Flask-Caching and Flask-Cache to provide caching functionality, which can improve the performance of web applications by reducing the number of requests made to the server

2.6 Data Base

MongoDB is a popular NoSQL document-oriented database that is designed for scalability, high availability, and performance. Unlike traditional SQL databases, MongoDB uses a flexible document data model, which allows for dynamic schemas and hierarchical data structures. MongoDB is an open-source database and is widely used in web applications, mobile apps, and other types of software

MongoDB is a powerful and flexible database that is well-suited for a wide range of applications, particularly those that require scalability and high performance. Its document data model and flexible schema make it easy to adapt to changing application requirements, while its integration with popular programming languages make it easy to use in a variety of software applications.