**System Architecture**

**for**

**Web scraper for latest news**

**3-Tier System Architecture**

|  |
| --- |
| 1. **Presentation tier** |
| This layer manages the user interface of the  Application. The frameworks and languages used for the user interface are HTML, CSS, Bootstrap. The homepage will display the latest news and wait for user interaction. The user requests will be sent to the 2nd(logic) tier and it will be handled in the following layers. The generated response will be received and displayed to the user. |

|  |
| --- |
| 1. **Application/Middle tier** |
| This layer fetches and processes the data to be displayed on the webpage. It also handles user requests like searching or filtering news. This application uses python-based web framework - Django for interacting with the presentation layer and data layer for data transfer and data storage respectively. |

|  |
| --- |
| 1. **Data tier** |
| This layer includes the data servers used for storing data and data access layer which provides APIs to the application layer for managing the data with minimum dependencies. This application uses the NoSQL database - MongoDB for data storage in the form of JSON objects. |

**Technologies Used**

Frontend : HTML5, CSS3, Bootstrap, JavaScript

Framework : Django

Database used : MongoDB (NoSQL)

HTML : Hyper Text Markup Language is used to create the main structure of a webpage

CSS : Cascading Style Sheets is used to define styles of HTML.

Bootstrap : It is a free and open source front end library for designing web applications and making it responsive.

Django : Django is a Python-based free and open-source web framework that follows the model-template-views architectural pattern.

MongoDB : It is a No-SQL database and uses Javascript Object Notation-like documents with schemas

**Libraries Used**

**Beautiful Soup** : It is a Python library for pulling data out of HTML and XML files. It provides a few simple methods and Pythonic idioms for navigating, searching, and modifying a parse tree.

## **Scikit-Learn** : It is the most useful and robust library for machine learning . It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistent interface in Python.

**Keras** : Keras is an [open-source](https://en.wikipedia.org/wiki/Open-source_software) [software](https://en.wikipedia.org/wiki/AI_software) library that provides a [Python](https://en.wikipedia.org/wiki/Python_(programming_language)) interface for [artificial neural networks](https://en.wikipedia.org/wiki/Artificial_neural_network). Keras acts as an interface for the [TensorFlow](https://en.wikipedia.org/wiki/TensorFlow) library.

**Testing Module**

Django’s ‘unittest’ library module will be used to simulate requests, insert test data, inspect your application’s output and generally to verify the code.

The test cases are to be written in test.py.

**Platforms**

**Version Control Management** : Version controlling helps us to manage the updated versions of the files in a software which is developed by multiple programmers. It is also helpful for peer review by other team members. Github will be used in this web application for version control.