

CHAPTER 6

SYSTEM IMPLEMENTATION

System Implementation is the stage where the theoretical design is converted into a working system, the new system may be totally new, replacing an existing manual, or automated system or it may be a major modification to an existing system. The system is implemented using Python and rainfall data set.

6.1 Programming Languages and Libraries Used

Python

Python is an interpreted high-level general-purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library.

Python is a multi-paradigm programming language. Object-oriented programming and structured programming are fully supported, and many of its features support functional programming and aspect-oriented programming (including by meta-programming and meta-objects (magic methods)). Many other paradigms are supported via extensions, including design by contract and logic programming.

Python uses dynamic typing and a combination of reference counting and a cycle-detecting garbage collector for memory management. It also features dynamic name resolution (late binding), which binds method and variable names during program execution.

Requests

Requests is a python module that allows us to send HTTP/1.1 requests extremely easily. There's no need to manually add query strings to your URLs, or to form-encode your PUT & POST data.

We use this module to make a API request to the data set provider and get the dataset as a response.

Scikit-learn

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.

Scikit-learn uses NumPy extensively for high-performance linear algebra and array operations. Furthermore, some core algorithms are written in Cython to improve performance. Support vector machines are implemented by a Cython wrapper; logistic regression and linear support vector machines by a similar wrapper. Scikit-learn integrates well with many other Python libraries, such as Matplotlib and plotly for plotting, NumPy for array vectorization, Pandas dataframes, SciPy, and many more.

Numpy

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. The ancestor of NumPy, Numeric, was originally created by Jim Hugunin with contributions from several other developers. In 2005, Travis Oliphant created NumPy by incorporating features of the competing Numarray into Numeric, with extensive modifications. NumPy is open-source software and has many contributors.

Pandas

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. It is free software released under the three-clause BSD license. The name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals. Its name is a play on the phrase "Python data analysis" itself.

Matplotlib

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK.

Pyqt5

PyQt5 is the latest version of a GUI widgets toolkit developed by Riverbank Computing. It is a Python interface for Qt, one of the most powerful, and popular cross-platform GUI library. PyQt5 is a blend of Python programming language and the Qt library. It is a popular C++ framework for writing GUI applications for all major desktop, mobile, and embedded platforms (supports Linux, Windows, MacOS, Android, iOS, Raspberry Pi, and more).

6.2 Methods for Flood Prediction

KNN (K-Nearest Neighbor)

Classification KNN is a nearest neighbor classification model where you can change both the distance matrix and the number of nearest neighbor. It stores training data, can use the model to compute the resubstitution prediction. This model can be convenient because training a classifier occurs in one step and classification in other steps.

Logistic Regression

Logistic Regression is a machine learning algorithm that predicts the probability of a categorical dependent variable. It is a statistical way of analyzing a set of data that comprises more than one independent variable that determines the outcome. The outcome is then measured with a dichotomous variable. The goal of this algorithm is to find the best model to describe the relationship between a dichotomous characteristic of interest and a set of independent variables.

SVM (Support Vector Machine)

It is a supervised learning algorithm which can used for binary classification or regression. It is a coordinate of individual observations. It is based on decision planes which defines decision boundaries. It also separated the set of objects having different class. a. This classifier is chosen as it is incredibly versatile in the number of different kernel functions that can be applied, and this model can yield a high predictability rate. SVM is one of the most popular and widely used clustering algorithms. It belongs to a group of generalized linear classifiers and is considered as an extension of the perceptron.

Decision Tree

Decision Trees (DTs) are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. A tree

can be seen as a piecewise constant approximation. Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal, but are also a popular tool in machine learning.

6.2 Procedure for Flood Prediction

Step 1: Start

Step 2: Input the region of which flood needs to be predicted from the drop down.

Step 3: Input the time of year for the prediction.

Step 4: Based on the given data, select the region and time of year from the dataset.

Step 5: Preprocess the data to remove any duplicates, null values etc.

Step 6: Display graphs and other statistics related to the selected information.

Step 7: Split the data into two sets, training and test set.

Step 8: Input the data into the different models.

Step 9: Get the predictions from the models and output the prediction along with the accuracy.

Step 10: End