



WEATHER FORECASTING APPLICATION

VISION STATEMENT DOCUMENT – STAGE 1

Course:

Data Science Product Development
(50742)

Group members:

Arsalan Shaikh - 23388
Ramsha Arif - 23395
Sarosh Aamir - 23390

BUSINESS OBJECTIVE

Weather forecasting is the application of scientific techniques and technology to predict the conditions of the atmosphere at a certain location and time. Weather Forecasting in old time is carried out by hand, using changes in barometric pressure, current weather conditions, and sky condition or cloud cover, weather forecasting now relies on computer-based models that take many atmospheric factors into account. accounting now relies on computer-based models that take many atmospheric factors into account.

Our main goal is to analyze how the different machine learning techniques will perform in the forecasting of weather combining time series with the neural network.

This project is divided into two parts, namely a research component and an application component, aiming to provide users with weather **predictions** using different machine learning models along with providing a good user experience with **interactive interface**.

UNDERLYING DATA SCIENCE PROBLEM

In this case, a software system can learn from data for improved analysis. Compared to traditional demand forecasting methods, machine learning helps to:

- Accelerates data processing speed
- Provides a more accurate forecast
- Automates forecast updates based on the recent data
- Analyzes more data
- Identifies hidden patterns in data
- Creates a robust system
- Increases adaptability to changes

USER STORY

User should be able to see following options in the web applications:

Main Page	Shows the current weather data of an auto-selected location. User can change the location.
Weather Details	On clicking detail button, user will see all the current values related to the selected location. Also, a visual and tabular representation of restricted past and future days.
Future Estimation	Using different advance machine learning techniques user will be able to see a comparative result of future weather.

TOOLS & TECHNIQUES

High-Level Approach:

The main goal is to create an application that can historical data of Pakistan and give optimum predictions.

In order to achieve this, the following pipeline is design:

- **Dataset-** will be hosted on S3 bucket (AWS).
- **Python-** will be the preferred language.
- **Amazon SageMaker-** will be used to prepare, build, train, and deploy machine learning models.
- **Boto3-** will be used to create, configure, and manage AWS services.
- **Flask(not fixed)-** might be used for local development and testing.
- **Amazon DynamoDB(not fixed)-** might be used to create, modify and retrieve data.
- **Docker Container** – will be used to manage entire life-cycle of application.
- **Other Libraries-** Pandas, Scikit-Learn, TensorFlow, Keras etc.
- **Frontend-** For frontend development JavaScript frame work with CSS and HTML will be used.

Further, Support Vector Machine (SVM), Artificial Neural Network (ANN) and a Time Series based Recurrent Neural Network(RNN) neural network or LSTM will be applied to make predictions and achieve high accuracy rate.

DATA SET

We have picked” Pakistan’s Integrated Surface Dataset” from National Centers of Environment Information [website](#).

Integrated Surface Data provides a long-term record of hourly, sub-hourly and synoptic weather observations from station networks around the world. Weather elements observed include. Online data are typically delayed for 24 hours.

Dataset consists of following files:

- **Elements:** sky conditions, visibility, weather type, dry bulb, wet bulb, and dew point temperatures, relative humidity, wind speed and direction, wind gusts, station and sea level pressure, altimeter, and precipitation.
- **Format:** .csv format.
- **Size:** ~600MB
- **Date Range:** 2011-2020