BIRZEIT UNIVERSITY

Faculty of Engineering and Technology

Electrical and Computer Engineering Department

ARTIFICIAL INTELLIGENCE

(ENCS3340)

Project Report

"Machine Learning Application for Weather Prediction in Ramallah"

Prepared by:

Hazar Michael **ID:**1201838

Rania Rimawi **ID:**1201179

Instructor: Dr. Adnan Yahya

Section: 1

Date: 29 January, 2024

Table of Contents

Table of figures

1. Introduction

In this project, we developed a weather forecasting application tailored for Ramallah, utilizing the power of machine learning. The data set comprising over 520 distinct weather records, our application employs two distinct models: a Decision Tree and an Artificial Neural Network (ANN).

These models are trained and tested on this dataset, enabling us to not only identify the more effective algorithm for this specific application but also to provide accurate weather predictions for the upcoming days in Ramallah. This project aims to demonstrate the practical application of machine learning algorithms in real-world scenarios, specifically in enhancing the accuracy and reliability of local weather forecasting.

2. Dataset description

The dataset utilized in this weather prediction project was sourced from Visual Crossing Weather Data Services, a provider of comprehensive historical, current, and forecast weather data. The dataset has a range of weather measurements recorded for the city of Ramallah over a period spanning from August 1, 2022, to January 1, 2024.

Accessed from: https://www.visualcrossing.com/weather/weather-data-services#

the dataset includes:

- Max Temperature (°C): The highest temperature recorded each day.
- Min Temperature (°C): The lowest temperature recorded each day.
- Average Temperature (°C): The daily mean temperature.
- Dew Point (°C): The temperature at which dew forms, indicating atmospheric moisture.
- Humidity (%): The percentage of moisture present in the air.
- Wind Speed (km/h): The speed of the wind measured in kilometers per hour.
- Wind Direction (°): The direction from which the wind originates, measured in degrees.
- Cloud Cover (%): The percentage of the sky covered by clouds.
- Date Components: Granular date information, including the year, month, and day.
- Icon: Categorical labels describing the weather conditions.

3. Methodology

3.1 Machine Learning Tools and Programming Languages

Our project uses Python due to its robust ecosystem of data analysis and machine learning libraries. We specifically employed the following tools:

- Pandas: For data manipulation and analysis.
- Scikit-learn: This one's a big deal in machine learning. We used it for splitting our data into two parts (one for training, one for testing), changing text labels into numbers, and making all the numbers play nice together.
- Keras: Served as our deep learning framework to build an Artificial Neural Network (ANN). We used the Sequential model to stack layers linearly and Dense for fully connected layers.
- Tkinter: Python's standard GUI toolkit was employed to develop a user-friendly interface for model interaction

3.2 Learning Algorithms

We selected two core machine learning algorithms to train, test and predict weather conditions:

1. Decision Tree Classifier (DT): A non-parametric supervised learning method used for classification The model was trained with the default settings for initial simplicity.

2. Artificial Neural Network (ANN): Our ANN consists of an input layer, two hidden layers with 64 neurons each, and a softmax output layer. The model utilizes the relu activation function in the hidden layers and is compiled with the Adam optimizer.

Both models were assessed based on their accuracy on the test set, which comprised 20% of the entire dataset. Additionally, a classification report was generated to evaluate precision, recall, and F1-score.

3.3 Data Preparation

The first task in preparing the data was to transform the date column into separate year, month, and day fields. This step aids the models in understanding patterns in the weather data. Additionally, the textual weather conditions represented by the 'icon' labels were converted into numeric codes using Label Encoding. This encoding process translated the weather conditions into the following codes: 'clear-day' as 0, 'partly-cloudy-day' as 1, and 'rain' as 2. This transformation was essential to ensure the dataset's compatibility with our machine learning algorithms, thereby preparing it for effective analysis and modeling.

4. Implementation

In our weather prediction project, the dataset was initially split into training and testing sets, with 80% of the data allocated for training and 20% for testing. Decision Tree Classifier and Artificial Neural Network (ANN), were trained on the training data.

The Decision Tree Classifier was trained to make weather predictions based on the provided features. Following that, the ANN model was constructed with multiple layers and was trained using a deep learning framework. After the training process was completed, both models were evaluated on the testing set to assess their accuracy, recall, precision and F1-score.

Finally, a prediction function was integrated into the graphical user interface (GUI). This function accepted user-provided input data, scaled it, and employed the previously trained models to generate predictions, facilitating accurate weather forecasts with user-friendly interaction.

5. Results

The User Interface of the project consists of labels and corresponding text input fields where users can provide input related to weather parameters. Additionally, there are four buttons: one to show Decision Tree evaluation results, another to display ANN evaluation results, a third for comparing model accuracies, and the last one for making weather predictions based on user input. Finally, there is an "Exit" button to close the application.

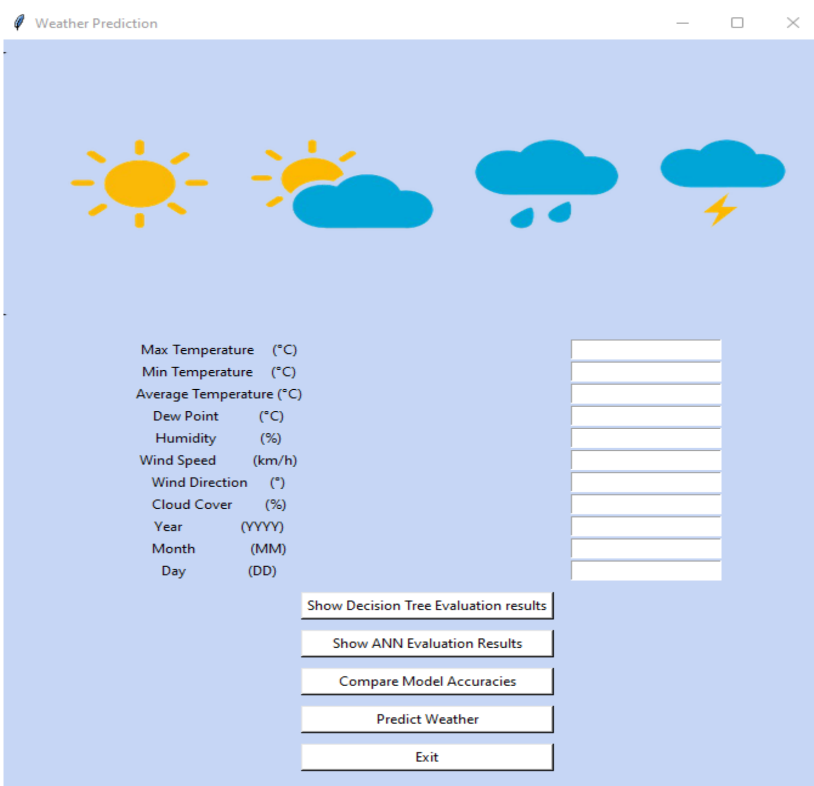

Figure 1: User Interface of the Weather Prediction Application for Ramallah

Figures 2 and 3 provide detailed evaluation reports for the Decision Tree and Artificial Neural Network models, which have been trained and tested on the Ramallah weather dataset. These reports include precision, recall, and F1-score metrics, offering a complete assessment of each model's performance.

Decision Tree Results

Accuracy: 0.8173

	precision	recall	f1-score
clear-day	1.000	1.000	1.000
partly-cloudy-day	0.848	0.765	0.804
rain	0.478	0.611	0.537
accuracy	0.817	0.817	0.817
macro avg	0.775	0.792	0.780
weighted avg	0.835	0.817	0.824

Figure 2: Decision tree evaluation results

ANN Results

Accuracy: 0.7981

	precision	recall	f1-score
clear-day	0.882	0.857	0.870
partly-cloudy-day	0.833	0.784	0.808
rain	0.591	0.722	0.650
accuracy	0.798	0.798	0.798
macro avg	0.769	0.788	0.776
weighted avg	0.808	0.798	0.801

Figure 3: Artificial Neural Network evaluation result

To improve the project's performance, a function that calculates and identifies the model with the higher accuracy was implemented. This function will be utilized to evaluate the predictions and determine which model provides more reliable forecasts.

Model Accuracy Comparison

Decision Tree Accuracy: 81.731%

ANN Accuracy: 79.808%

The model with the highest accuracy is Decision Tree (81.731%).

Figure 4: Model Accuracy Comparison

Furthermore, to assess the proximity of our predictions to actual conditions, we have integrated data from forecasts available on the web page for upcoming days. This approach ensures that our predictions closely align with real-world scenarios by using information sourced from the webpage as user input.

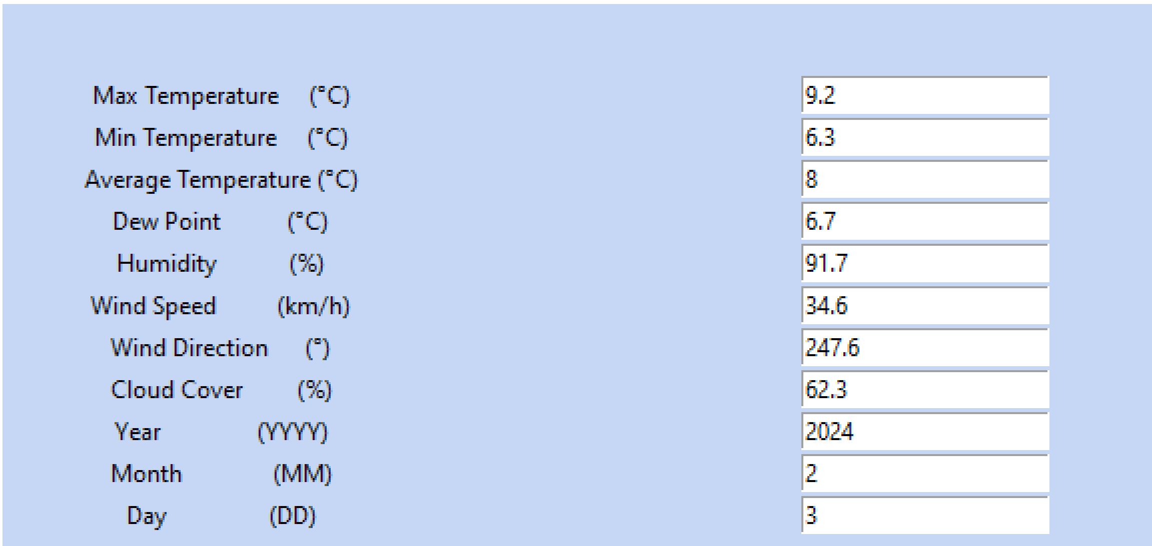

Figure 5: User input for predicting weather

In the figure below, it is evident that both models produce predictions that closely match the values provided by the website, indicating the accuracy of our models in forecasting weather conditions.

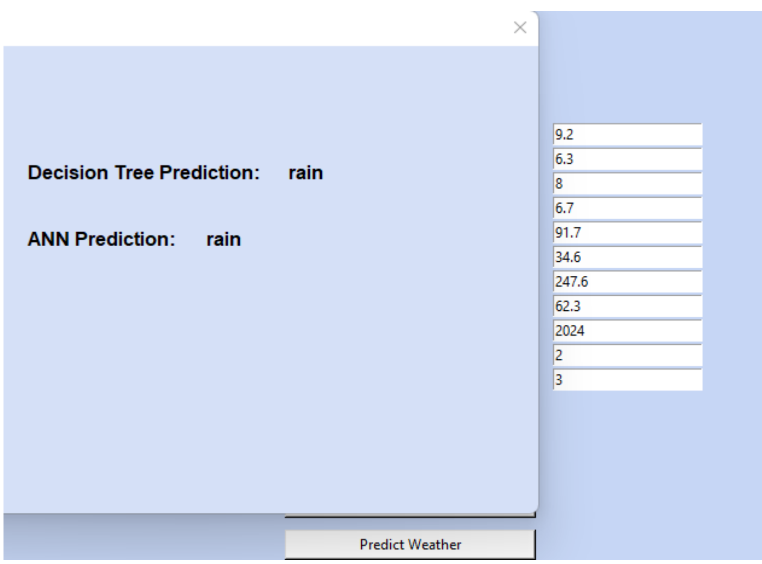

Figure 7: Model Predictions

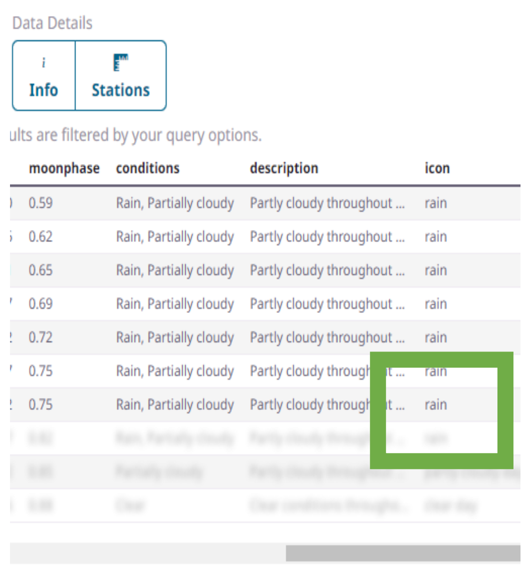

Figure 6: Website Predictions

6. Conclusion

Our project's training and testing phases were key to understanding the effectiveness of our models. During training, we fine-tuned our Decision Tree and Neural Network with historical weather data from Ramallah, aiming to capture the difficult patterns. Testing, on the other hand, served as a reality check, revealing how well our models could generalize to new, unseen data.

For the prediction phase, the Decision Tree model achieved better accuracy for predicting Ramallah's weather with about 82% accuracy, while the Neural Network scored around 80%. Typically, complex models like Neural Networks perform better, but the Decision Tree excelled, particularly in predicting clear days. The smaller dataset size might have limited the Neural Network, which tends to require more data to avoid overfitting.

We recognized that the variability in outcomes across different executions is a normal occurrence. This can happen due to random factors like how the data is split, the initial random weights of models, and the randomness in training algorithms.