

Coats ML Assignment

Group number: G6

Santhoshkumar. S

Part - A **Research**

Title of Paper: Classification of weather conditions based on supervised learning

Paper Contribution:

Machine learning algorithms are used for weather forecasting to achieve better results than traditional methods which rely on physical models such as temperature, humidity, wind, and cloud cover. The algorithms used and compared are,

- Artificial Neural Networks (ANN)
- Support Vector Machine (SVM)
- K-Nearest Neighbours (KNN)
- Random Forest (RF)

These algorithms are used to find hidden patterns and relationships between inputs and outputs, to improve the accuracy of weather forecasting, and also compared with weatherstack API.

Data:

Weatherstack API has been used to obtain 13 years of historical weather data on selected four cities around Sweden. Weatherstack is chosen because it has consistently displayed a high level of reliability and accuracy.

Data Preprocessing:

- Removal of Duplicates from dataset.
- Optimal hyperparameters for each model are found using k-fold cross validation with k=5. Model is trained using 4 folds and tested on the last one, the best results obtained from each fold are chosen to determine optimal hyperparameters. This process is done for all the models to ensure optimal performance.

Machine Learning Activity:

In ANN, to avoid overfitting, two critical metrics are used- epoch loss and epoch accuracy. These are plotted along with loss for the model and validation accuracy along the number of epochs. When epoch loss meets with model loss or epoch accuracy meets with validation accuracy, model training is stopped to avoid overfitting. This is an

early stopping rule for ANN to avoid overfitting used in this research paper. Other models are implemented using scikit-learn library.

Result Analysis with metrics:

The four models' accuracy is compared with each other after training it using 13 years of historical weather data from four different cities. ANN and RF proved to be better ones among these four with 97% and 96.69% accuracy while SVM and KNN are on the lower end with 93% and 77.97%. Similarly, ANN and RF have higher F1 scores than SVM and KNN, indicating strong balance between precision and recall.

Metrics:

Accuracy, Precision, Recall, F1 score, and Matthew's Correlation Coefficient (MCC) are the five metrics used to evaluate the models in this paper.

Significance:

There are no freely available research papers (except this one) currently on the web regarding Machine learning algorithms used for weather forecasting with detailed findings. This paper tells us the better machine learning algorithms which allows us to try other algorithms and compare all to achieve even more significant results.

Exploratory Data Analysis / Visualization:

A histogram for 'weather conditions' is plotted from which it can be observed that 'weather conditions' is a right skewed distribution.

Drawbacks:

Dealing with weather forecasting comes with inconsistencies and so further research must be done to tell the limitations.

Source: <https://www.diva-portal.org/smash/get/diva2:1765752/FULLTEXT01.pdf>