WEATHER CLASSIFICATION BASED ON HO CHI MINH CITY WEATHER DATA

Phan Thanh Hải[1]
18520705

Lê Quang Thắng[2]
16521099

Trương Tấn Sang[3]
18521336

Department of Computer Science, University of Information Technology

CS114.L11: Machine Learning

Dr. Nguyễn Lưu Thuỳ Ngân, Dr. Dương Ngọc Hảo

January 1, 2022

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
UIT

KHOA HỌC MÁY TÍNH

Table Of Contents

Chapter 1.
Introduction

1. Introduction

Goal. Develop a weather prediction model by utilizing HCM's 2020 weather data.

Definition.

Input. Time, Temperature, Humidity, Wind, Visibility, etc.

Output. Weather status (e.g., Clear, Partially cloudy, etc.).

Application. Providing information support for efficient planning in:

Meeting Schedules

Event Organization

Sports Activities, etc.

Chapter 2. Data Preparation

2. Data preparation

Date time	Temp	Wind Speed	Visibility	Cloud Cover	Relative Humidity	Conditions
1/1/2020 0:00	26	7.6	10	27	67.03	Partially cloudy
1/1/2020 1:00	25.9	9.4	10.2	25.4	73.53	Partially cloudy
1/1/2020 2:00	25	10.1	8	27	71.96	Partially cloudy

8784 samples

7 attributes

Crawl data from

Visual Crossing

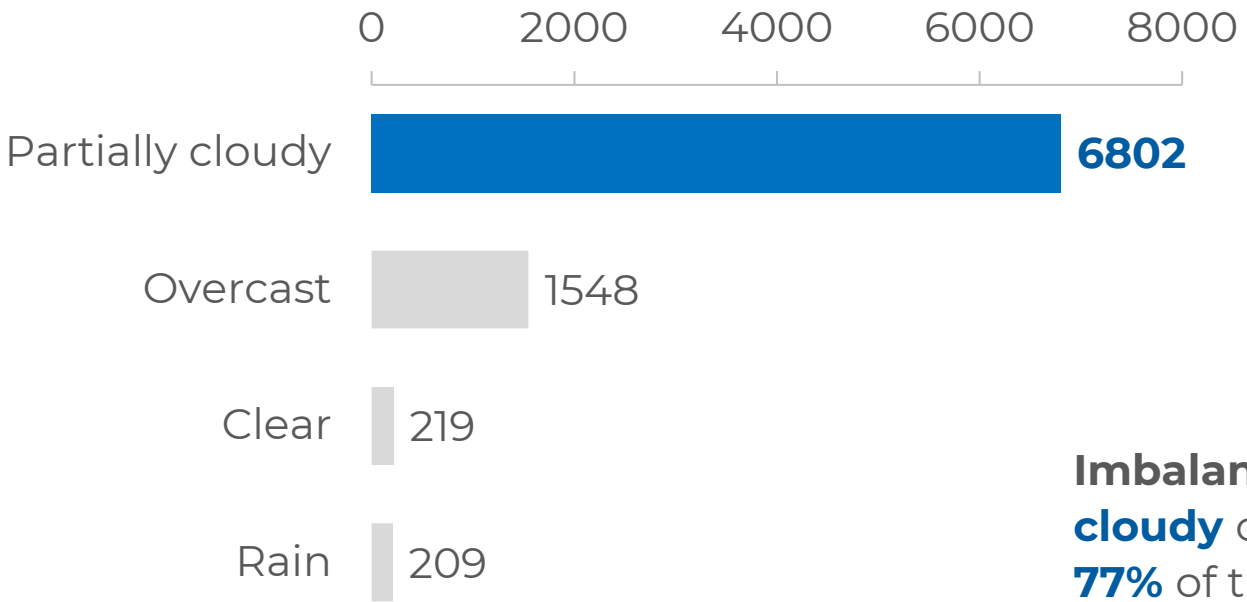
80% training data

compared to 20% test data

2. Data preparation

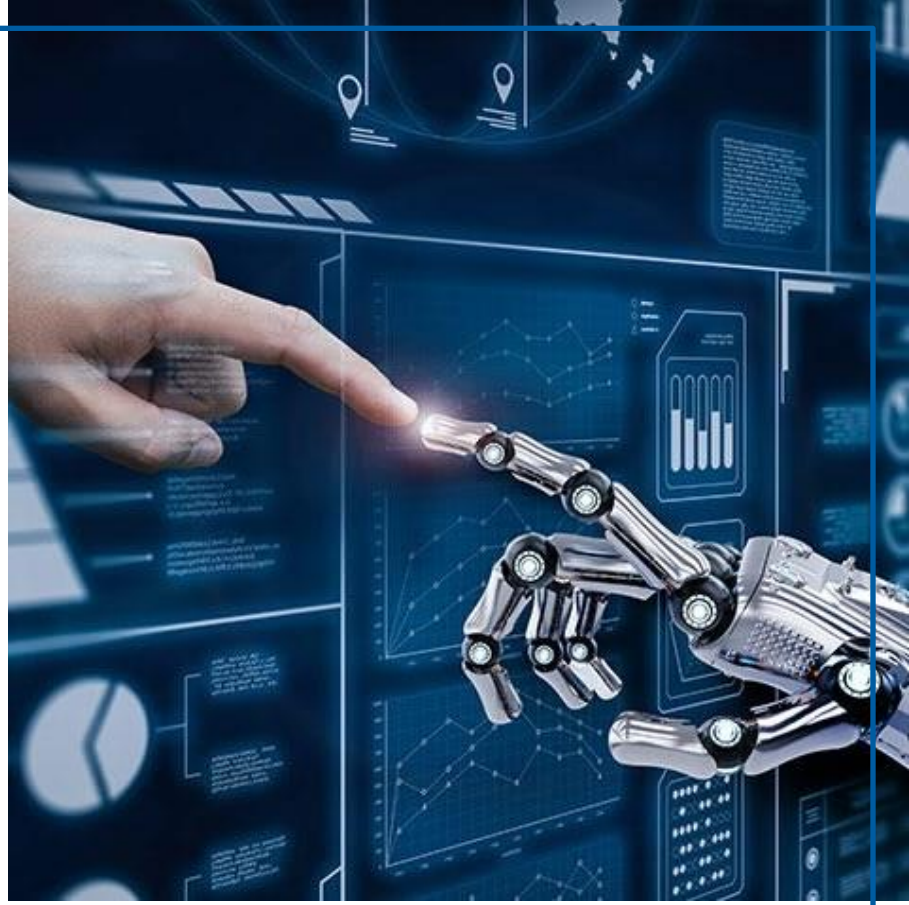
Distribution of Values in the **Conditions** Attribute

Is your dataset **imbalanced** or not?



Imbalanced dataset. **Partially cloudy** comprises slightly over **77%** of the values.

Chapter 3. Stages of building ML model



3. Stages of building ML model

Data Cleaning:
Exclude 8 samples with NULL values.

Data Discretization:
Date time and Conditions attributes.

Data Scaling:
Standardization.

3. Stages of building ML model

Attributes	Correlation Coeff
Hour	– 0.029
Day	– 0.0022
Month	– 0.086
Year	**0**
Temperature	– 0.13
Wind Speed	– 0.044
Visibility	0.012
Cloud Cover	– 0.36
Relative Humidity	0.17

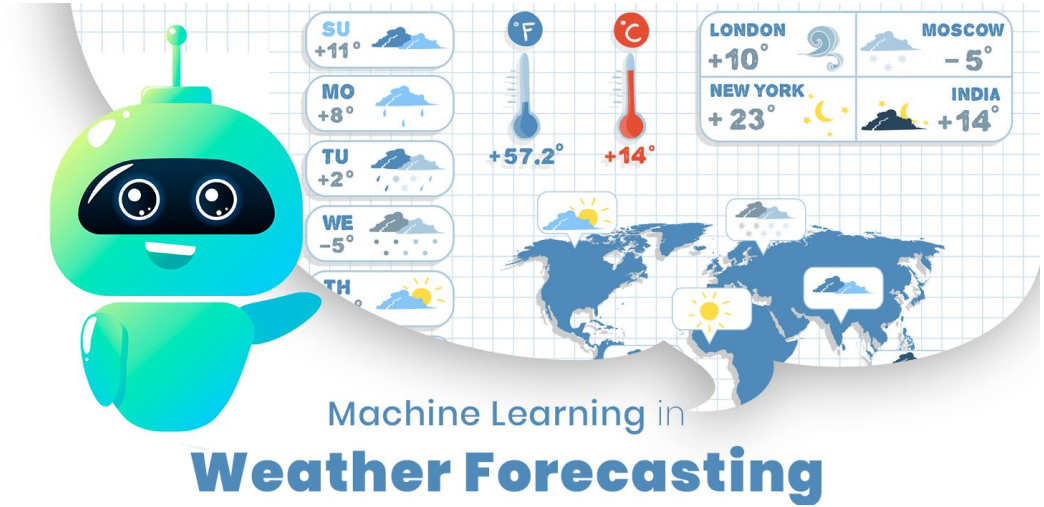
3. Stages of building ML model

Mô hình	Macro-average			Weighted-average		
	P	R	F1	P	R	F1
Decision Tree	85.21	89.60	**86.97**	97.83	97.27	**97.51**
Random Forest	92.71	88.82	**90.44**	98.39	98.52	**98.42**
Logistic Regression	73.03	74.98	**73.99**	95.33	97.61	**96.46**
SVM	72.97	73.85	**73.39**	95.22	97.49	**96.34**
K-nearest Neighbor	69.92	67.22	**68.39**	93.58	93.51	**93.52**

04 CONCLUSION

4. Conclusion

- What is **Achieved**?
- What is **Not Achieved**?
- **Future Development** Direction



- HẾT -

GIẢI ĐÁP THẮC MẮC
HỎI XOÁY ĐÁP XOAY