

WEATHER CLASSIFICATION BASED ON HO CHI MINH CITY WEATHER DATA

Phan Thanh Hải¹
18520705

Lê Quang Thắng²
16521099

Trương Tấn Sang³
18521336

Department of Computer Science, University of Information Technology
CS114.L11: Machine Learning

Dr. Nguyễn Lưu Thùy Ngân, Dr. Dương Ngọc Hào
January 1, 2022



TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ THÔNG TIN
UIT



KHOA HỌC MÁY TÍNH

Table Of Contents

1 INTRODUCTION

2 DATA ANALYSIS & PREPROCESSING

3 BUILDING MACHINE LEARNING MODELS

4 CONCLUSION



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 Analysis & Preprocessing



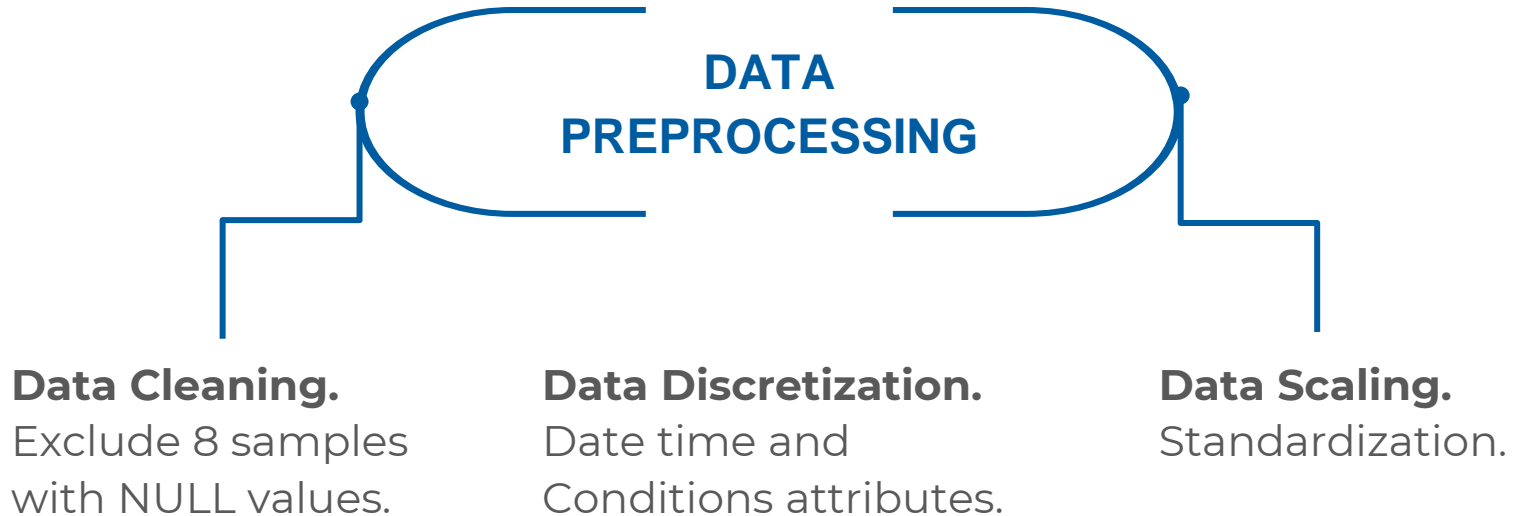
2. Data Analysis & Preprocessing

Date time	Temp	Wind Speed	Visibility	Cloud Cover	Relative Humidity	Conditions	— 8784 samples —
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	
7 attributes							

Crawl data from
Visual Crossing

80% training data
compared to 20% test data

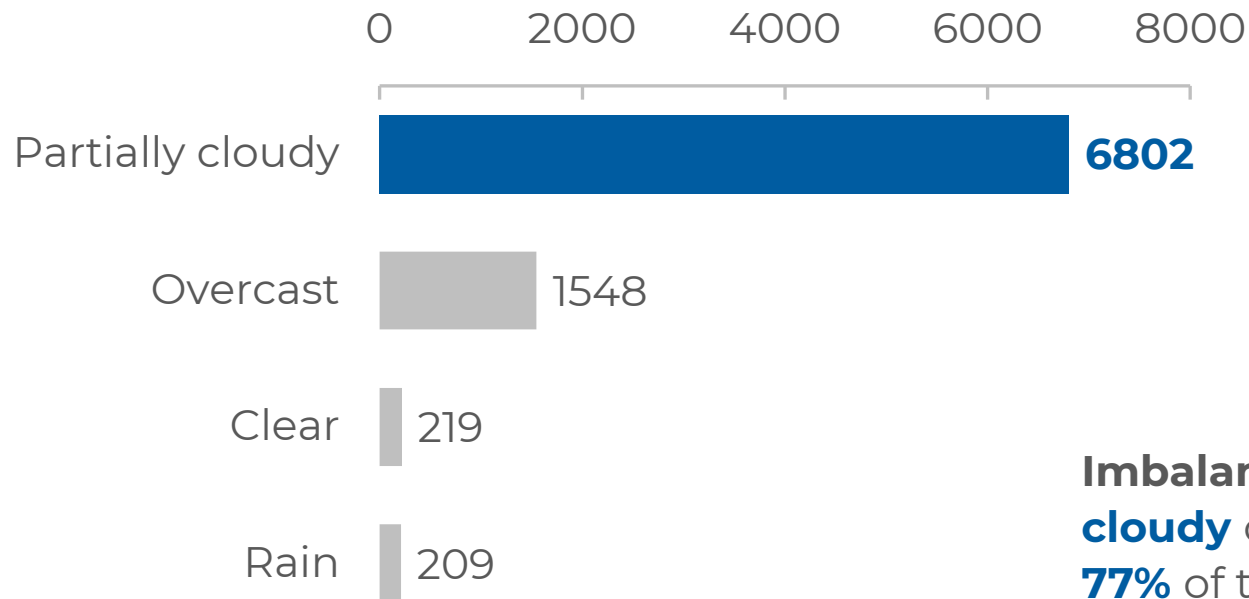
2. Data Analysis & Preprocessing



2. Data Analysis & Preprocessing

Distribution of Values in the **Conditions** Attribute

Is the dataset **imbalanced** or not?



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

2. Data Analysis & Preprocessing

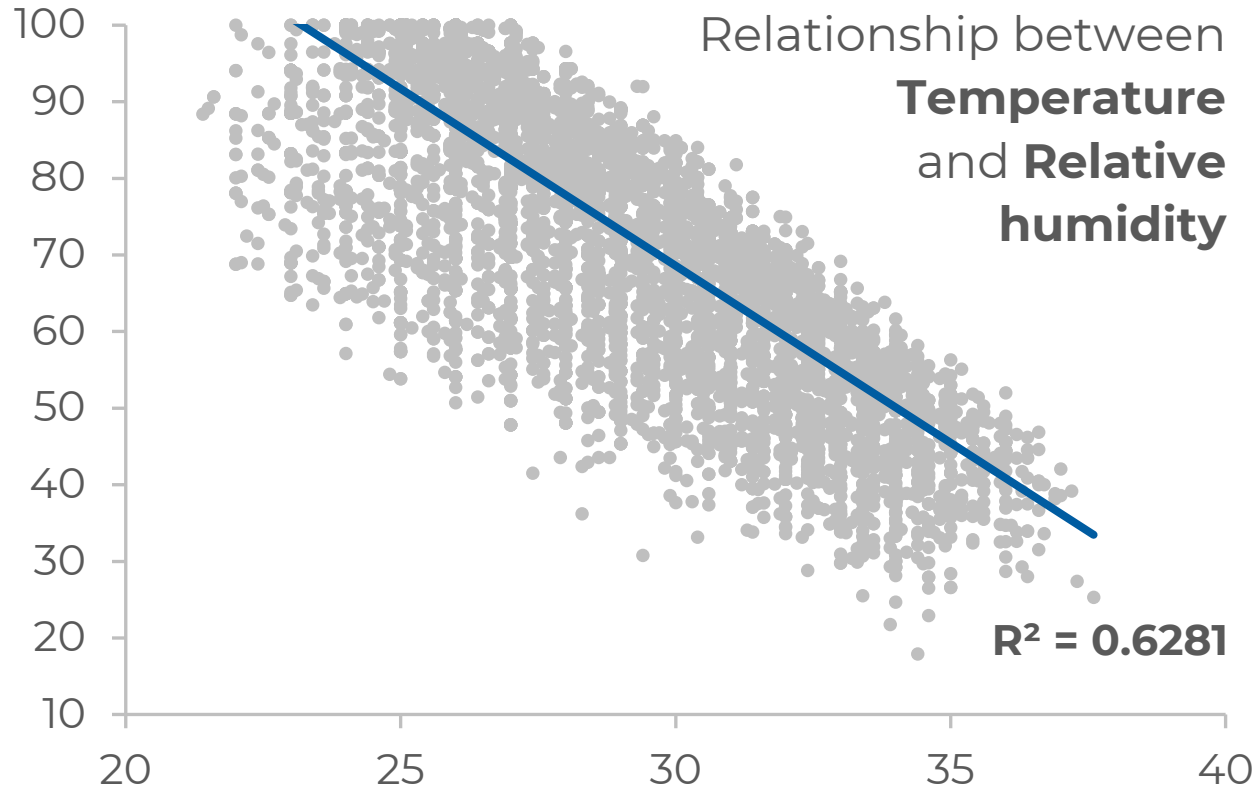
Correlation Matrix

Wind Speed	0.28							
Visibility	0.34	0.26						
Cloud Cover	0.12	-0.02	-0.13					
Humidity	-0.79	-0.37	-0.41	0.07				
Conditions	-0.13	-0.04	0.01	-0.36	0.17			
Month	-0.18	-0.21	-0.2	0.42	0.34	-0.09		
Day	0.01	-0.02	0.01	0.05	0.01	0.0	0.01	
Hour	0.29	0.28	0.14	0.1	0.27	-0.03	0.0	0.0
	Temp	Wind Speed	Visiblity	Cloud Cover	Humi dity	Condi tions	Month	Day

There is a strong **inverse relationship** between temperature and relative humidity.

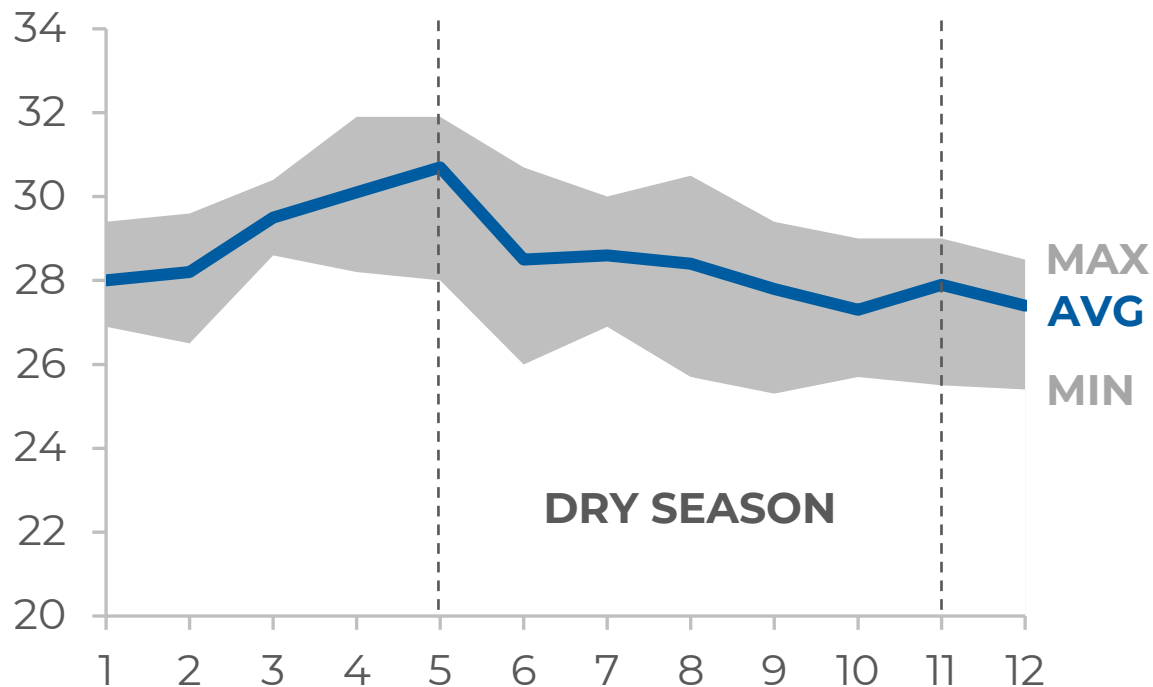
Focus on **Temperature**, **Cloud Cover** and **Relative Humidity** for predicting Conditions.

2. Data Analysis & Preprocessing



2. Data Analysis & Preprocessing

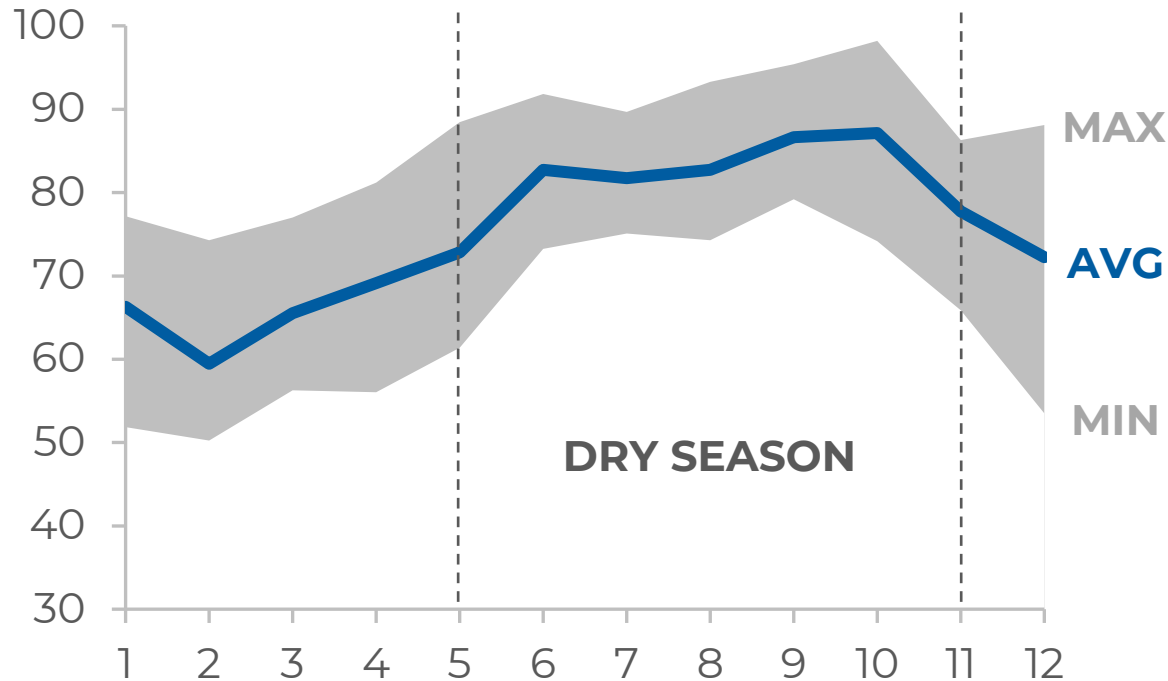
HCMC's 2020 Average Monthly Temperature



The city has a **tropical savanna climate** with **high** average temperatures throughout the year.

2. Data Analysis & Preprocessing

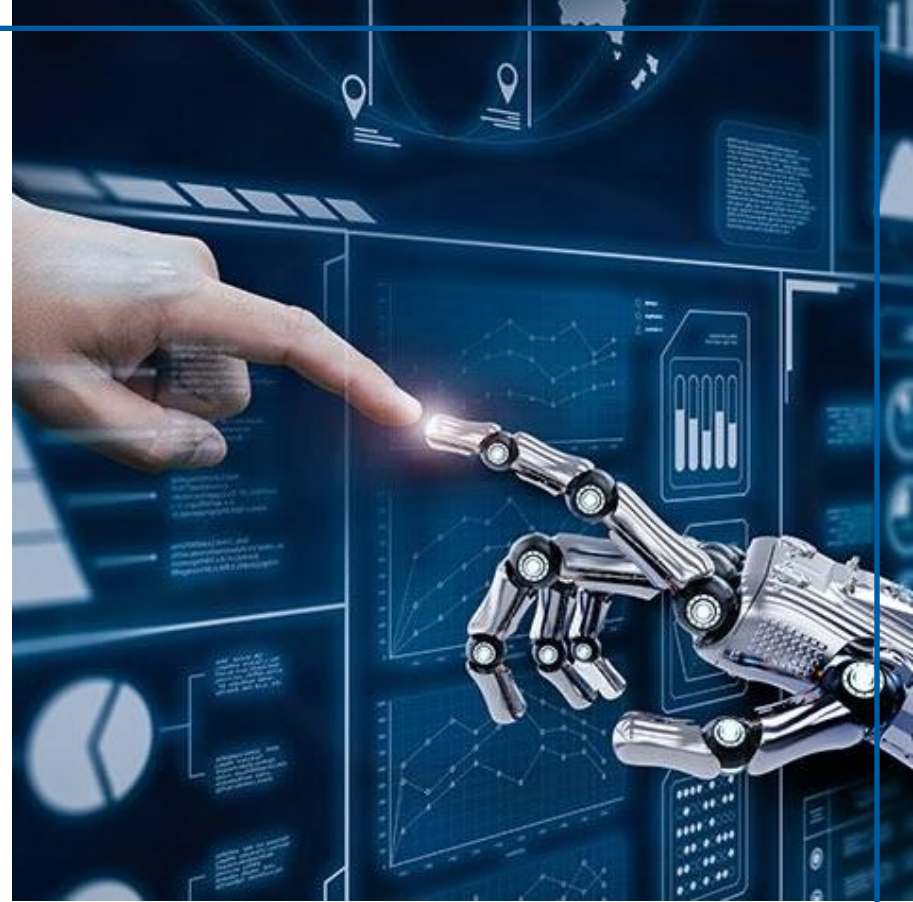
HCMC's 2020 Average Monthly **Relative Humidity**



The city has a **tropical savanna climate** with **high** moisture content of the air.

Chapter 3.

Building Machine Learning Models



3. Building Machine Learning Models

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

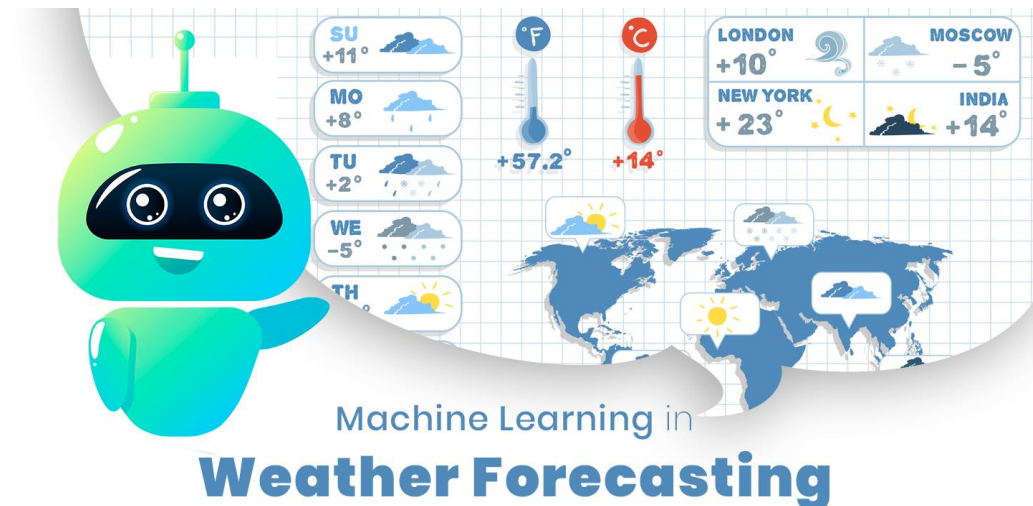


04

CONCLUSION

4. Conclusion

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



- THE END -

Q and A

Questions are
Needed, Really?

