Group name:
404 NOT FOUND

SHORT REPORT

DISCRETE STRUCTURES FOR COMPUTING ASSIGNMENT
Semester: 202

TEAM MEMBERS

NAME	ID	CONTRIBUTION	PERCENTAGE
Nguyễn Nho Gia Phúc (Leader)	2052214	• Team management (meeting). • LaTeX 1, 2, 3, 4, 6.8, 6.10, 7, 8.	20%
Nguyễn Văn Quốc Chương	1950004	• Time Series, Linear Regression, Decision Tree, SVM Models. • Model Analysis • Appendix and theory for Linear Regression. • LaTeX chapter 6.1 and 11.	20%
Thái Tài	2052246	• SIR and Time Series Theory. • LaTeX chapter 6. • SIR Model. • Short report.	20%
Nguyễn Trọng Chuẩn	2052046	• Analysis: Time-based, Location-based, and other analyses. • Developed analysis strategies.	20%
Vũ Đồng Tuệ Quyên	2052234	• Developed analysis strategies. • LaTeX chapter 5. • Distribution and Vaccines theory parts.	20%

WEEK 1 + 2

1/6	**All group members: Learned tools (python, pandas, numpy...)**
5/6	All group members: Collected and exchanged data
8/6	• Tài: Time series • Phúc: LaTeX 1 • Chương: Learned models • Chuẩn: Learned more tools • Quyên: History of Vaccine
12/6	• Tài: Time series decomposition • Phúc: LaTeX 2 • Chương: ARIMA model • Chuẩn: Learned more tools + Developed analysis strategies • Quyên: Developed analysis strategies

WEEK 3 + 4

15/6	• **Tài: ARIMA** • **Phúc: LaTeX 3** • **Chương: Learned models** • **Chuẩn: Time-based code** • **Quyên: Developed analysis strategies**
19/6	• Tài: Holt-Winters • Phúc: LaTeX 4 • Chương: Holt-Winters • Chuẩn: Time-based code • Quyên: Distribution theory
22/6	• Tài: SIR theory • Phúc: LaTeX 6.8 • Chương: Learned model • Chuẩn: Location-based code • Quyên: Distribution theory
26/6	• Tài: SIR compartmental models • Phúc: LaTeX 6.8 • Chương: Supervised learning model • Chuẩn: Location-based code • Quyên: Developed additional analysis strategies

WEEK 5 + 6

29/6	• **Tài: Wrote LaTeX 6.6, 6.7, 6.9, 6.11 ,6.12** • **Phúc: LaTeX 6.10** • **Chương: Appendix B** • **Chuẩn: Appendix A** • **Quyên: Developed additional analysis strategies**
3/7	• Tài: Wrote LaTeX 6.2, 6.3, 6.5, 6.13 • Phúc: LaTeX 6.10 • Chương: Appendix B • Chuẩn: Appendix A • Quyên: LaTeX 5
6/7	• Tài: short report • Phúc: LaTeX 9 • Chương: Appendix B, Model Analysis • Chuẩn: Appendix A • Quyên: LaTeX 5
10/7	• All group members: Grammatical check and minor bug fixes. • Phúc: report submission with links and supporting files.

1. INTRODUCTION

- Introduce the COVID-19 situation

2. TOOLS AND ENVIRONMENTS

-PYTHON
-PANDAS
-NUMPY
-MATPLOTLIB
-SKLEARN
-COLAB
-PMDARIMA
-SEABORN

3. MACHINE LEARNING

-SUPERVISED LEARNING

-UNSUPERVISED LEARNING

- REINFORCEMENT LEARNING

- ENSEMBLE LEARNING

- DEEP LEARNING

4. DATA COLLECTION

-DATA COLLECTION

-DATA PRE-PROCESSING

-DATA DESCRIPTION

5. ANALYSIS

5.1 Overview of vaccine

- History of vaccine
- Variety of vaccine
- Income level versus vaccine

5. ANALYSIS

5.2 Vaccine Analysis

- Vaccine prevents the spread of the virus

- Divided the data into 4 groups and 2 periods (July – Dec 2020) and (Jan – June 2021):

up-up, down-down, up-down, down-up

-Vaccine affects the deaths and age groups

=> Most elderly and middle-aged have more vaccine than adolescents. This priority is rooted in the fact of background disease.

5. ANALYSIS

5.3 Other analysis

- Delta variant of Coronavirus
- Covid situation in Vietnam
- Correlation of matrix

5. ANALYSIS

5.4 Distribution theory

-Definition + Application

- Normal distribution
- Bimodal distribution
- Poisson distribution
- Gamma distribution
- Exponential distribution

5. ANALYSIS

5.5 Distribution Analysis

-Distribution of new infections

⇒ infections tend to level off or decrease in the second stage in countries with high vaccinations rate.

-Distribution of vaccinated age

⇒ The median is 60-69 aged group

-Delta variant in new infections

6. PREDICTION MODELS

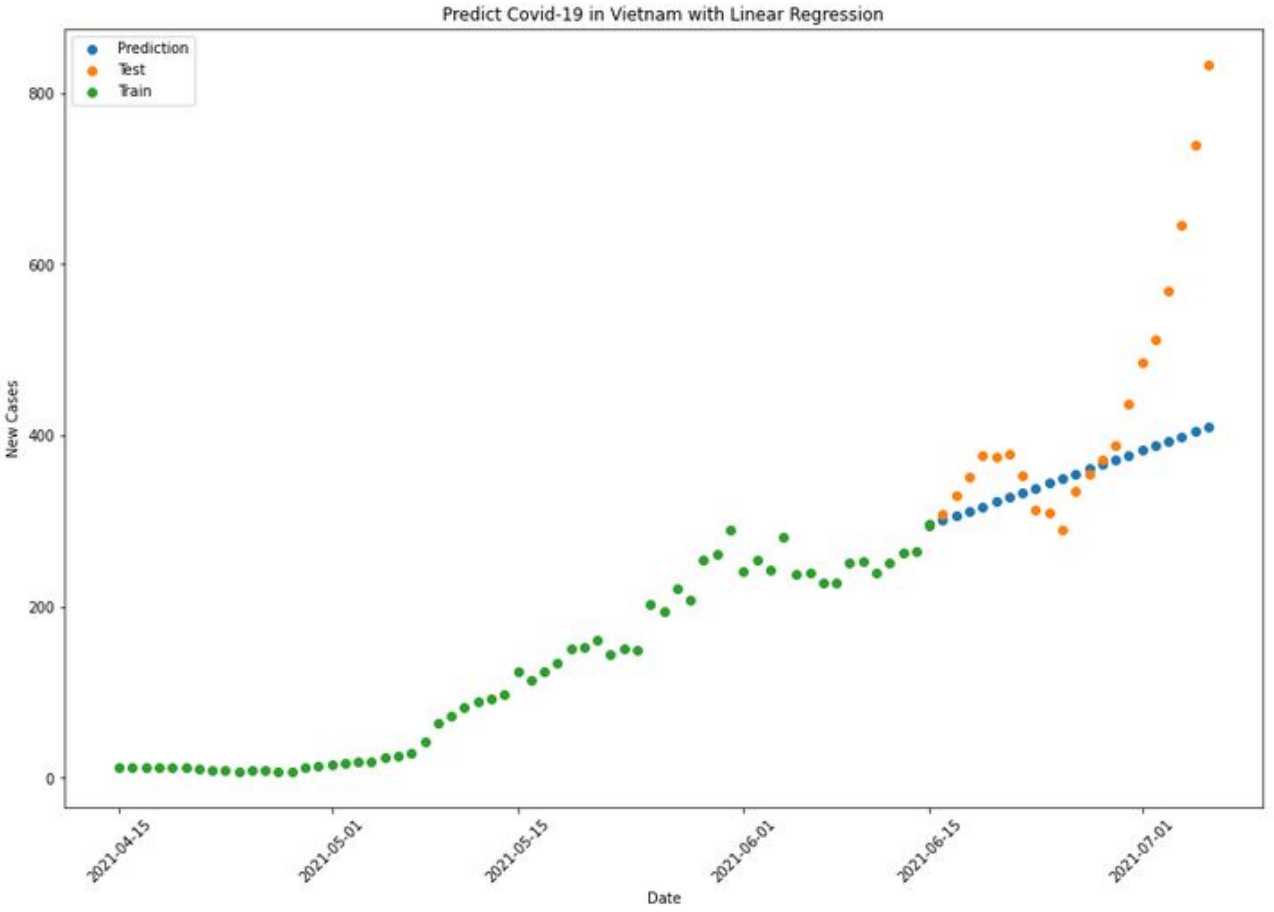
6.1 – 6.5: Supervised Learning model

6.6 – 6.10: Time Series model

6.12 - 6.14: SIR model

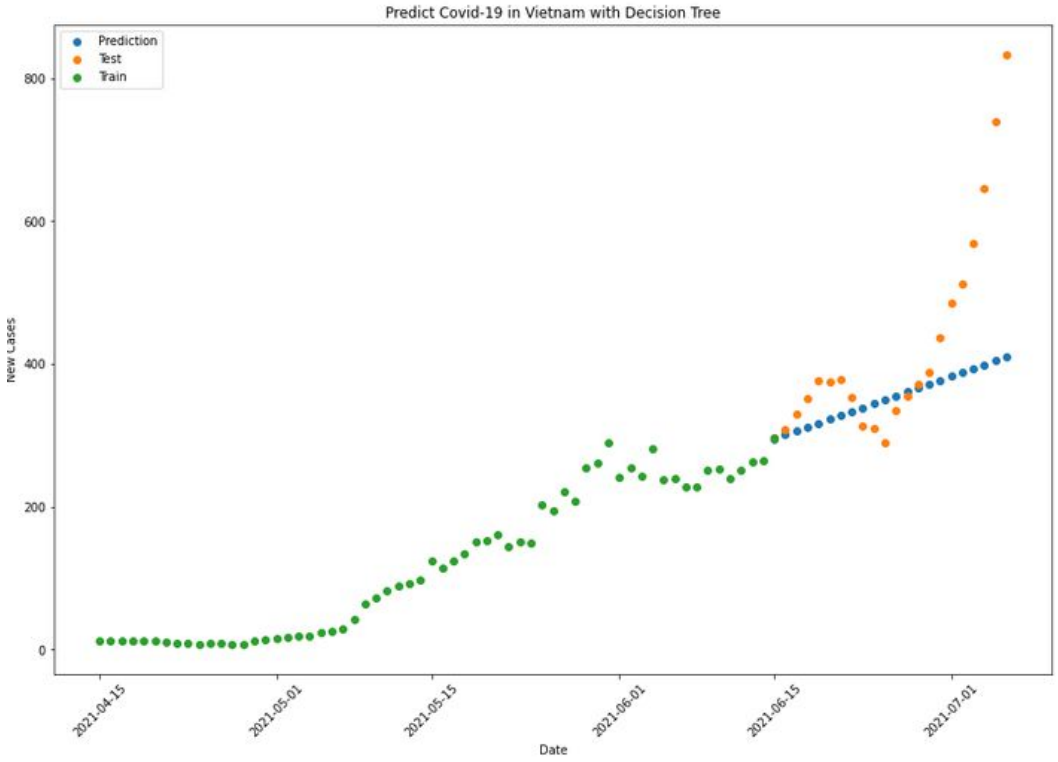
6. PREDICTION MODELS

LINEAR REGRESSION



6. PREDICTION MODELS

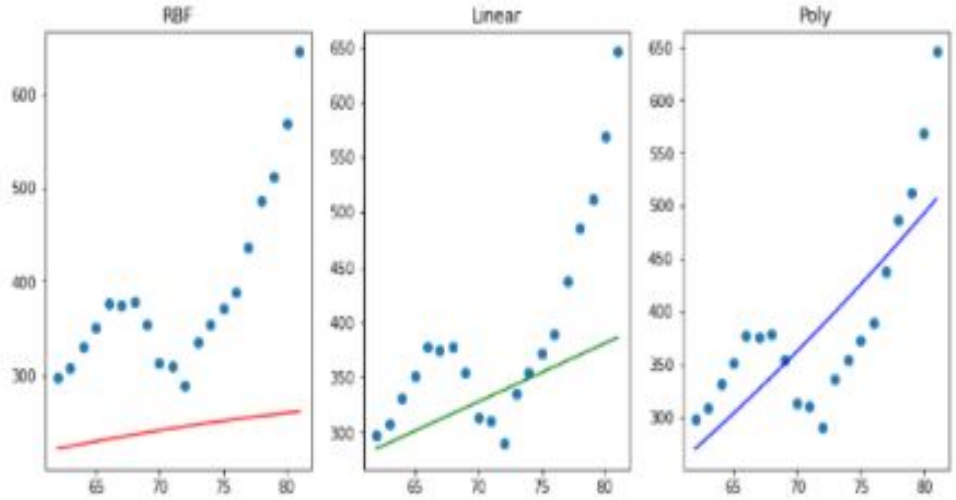
DECISION TREE

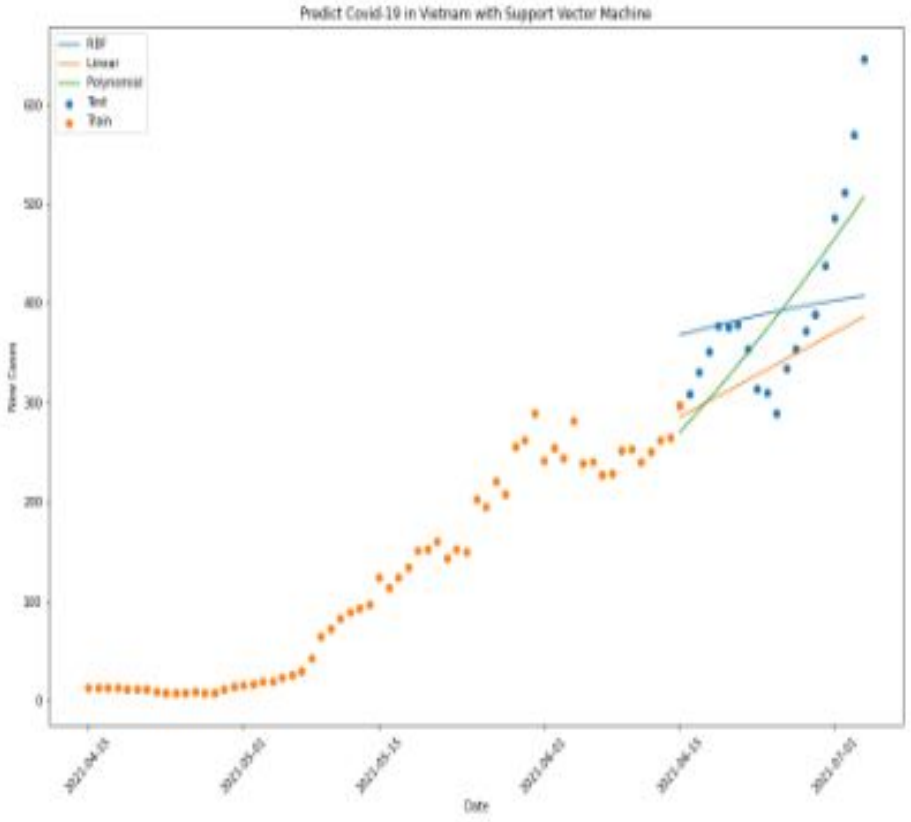


Has the same result like Linear Regression

6. PREDICTION MODELS

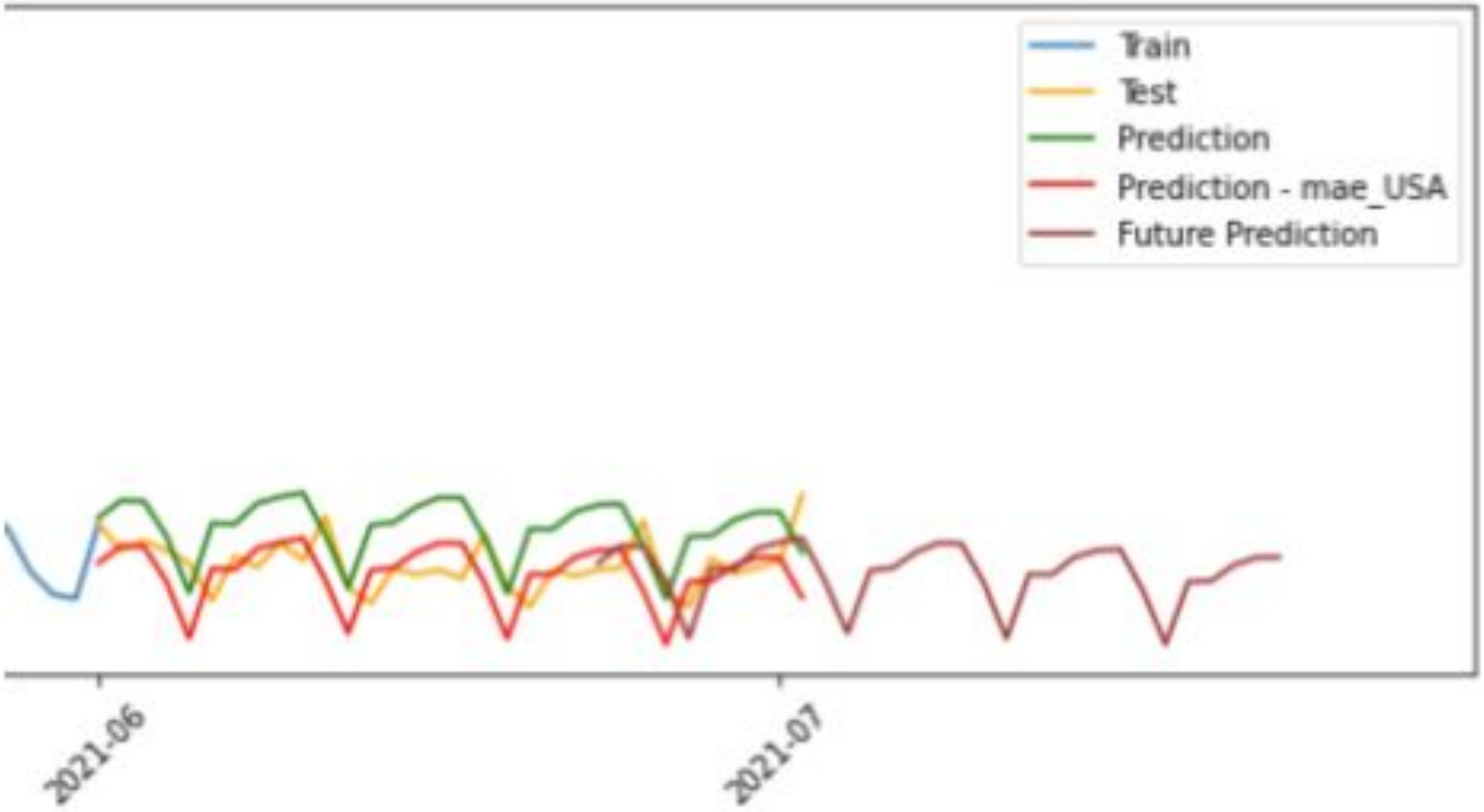
SUPPORT VECTOR MACHINE





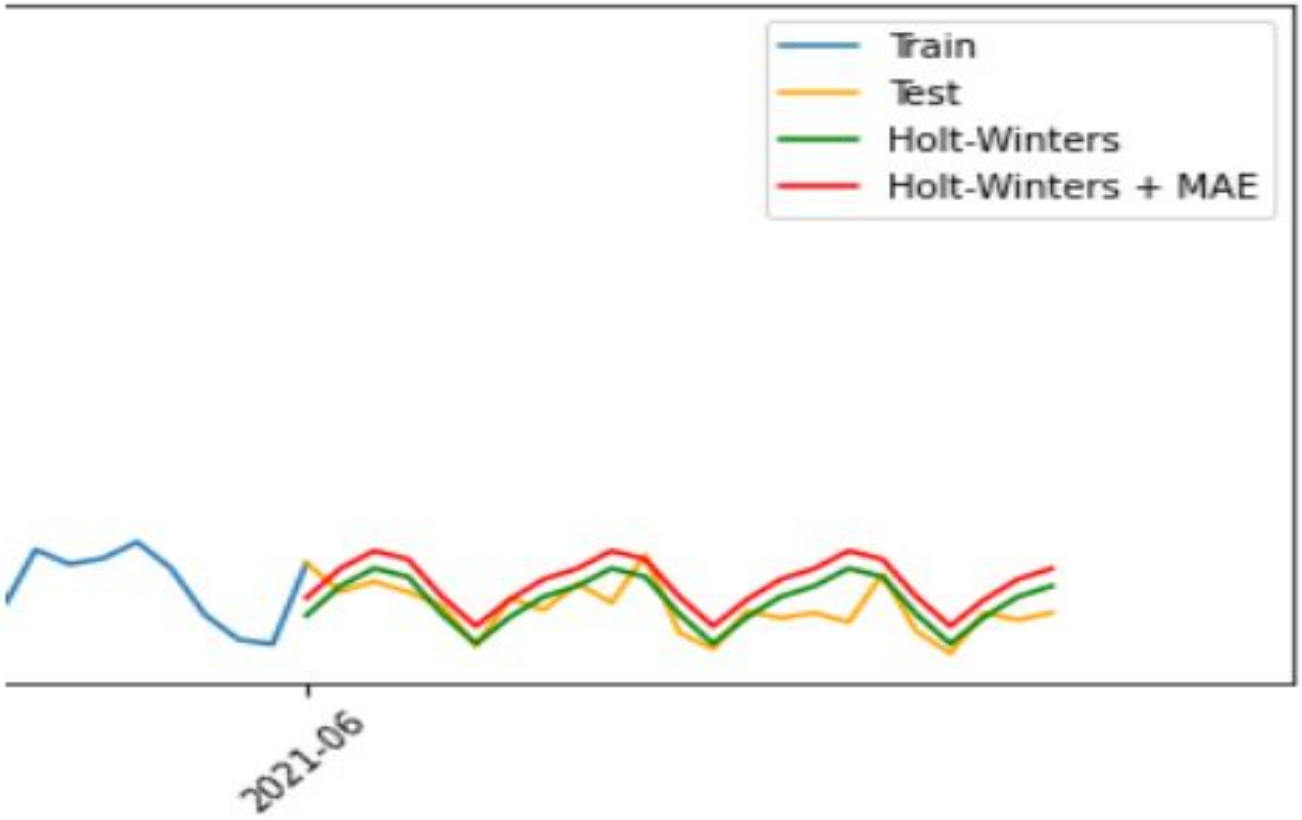
6. PREDICTION MODELS

ARIMA



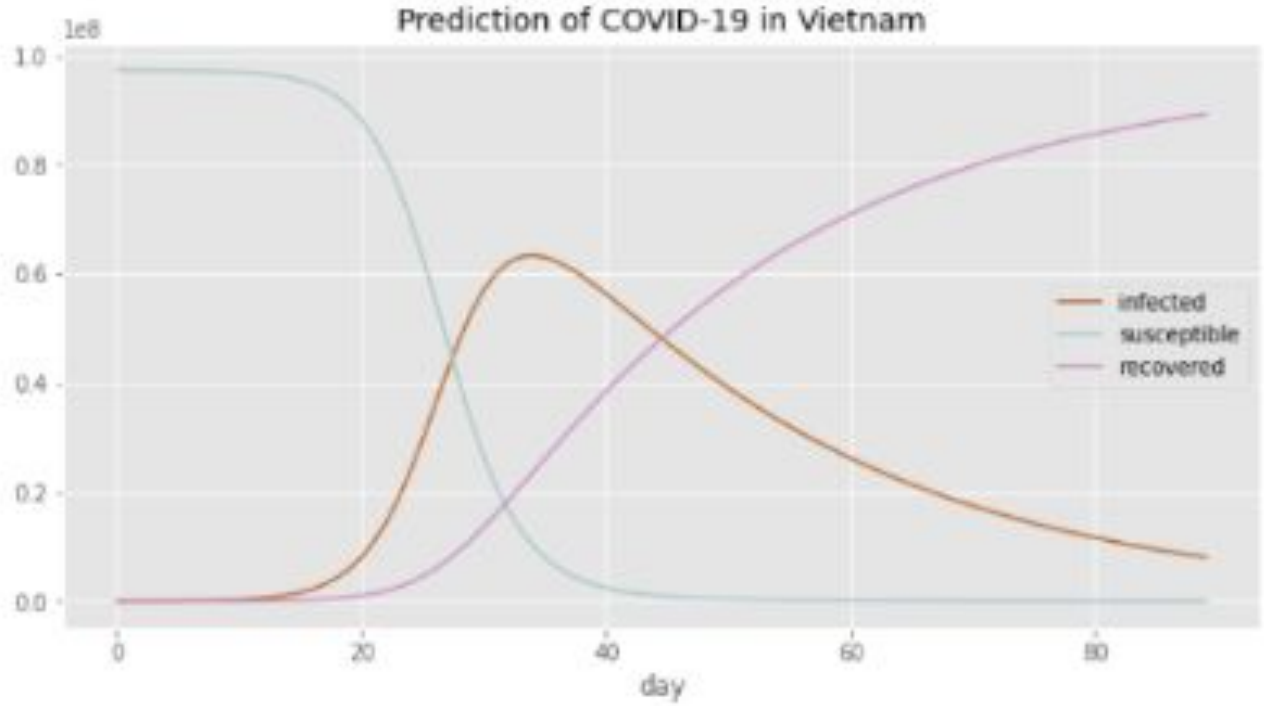
6. PREDICTION MODELS

HOLT-WINTERS



6. PREDICTION MODELS

SIR MODEL



7. MODEL EVALUATION

Testing: Using MAE, MSE, R^2, RMSE.

=> The lower the error, the better the model.

7. MODEL EVALUATION

Linear + Decision tree: Decision tree relies on probability, performs similarly to linear regression because the data has little features.

7. MODEL EVALUATION

SVM is not appropriate for this data. (gamma vs. C)

It is more useful in classification problems.

7. MODEL EVALUATION

SIR:

- Ignores some parameters that may affect the result.
- Utopian hypothesis: 1/ population never changed

  2/ recovered get the immunity

7. MODEL EVALUATION

Time series: HW and ARIMA - Good for short-term prediction but not long-term due to minimal data input.

7. MODEL EVALUATION

- Cross-country model: usable if the trends are similar.
- Based on observations and experience, we can alter the MAE for better result.
- ARIMA is one day ahead of the actual, but the effect is not abysmal.

8. CONCLUSION

INFERENCE FROM OUR REPORT

9 & 10. LINK CODE

2 LINKS ANALYSIS:

- COUNTRY-BASED
- -TIME-BASED

3 LINKS MODELS:

- SUPERVISED LEARNING MODEL
- TIME SERIES MODEL
- SIR MODEL

5 LINKS OTHER MODELS AND TECHNIQUES

- LINEAR REGRESSION
- LOGISTIC REGRESSION
- NEURON NETWORK
- K-MEANS
- PRINCIPAL COMPONENT ANALYSIS

11. APPENDIX

A. DATA PREPROCESSING
B. BASIC ALGORITHMS

11. APPENDIX

A. DATA PREPROCESSING

- Definition
- Data cleansing
- Data integration
- Data transformation
- Data reduction

11. APPENDIX

B. BASIC ALGORITHMS

- Supervised Learning
  + Logistic regression
- Unsupervised Learning
  + K-Means clustering

12. REFERENCES

MANY REFERENCES ARE LISTED

THE END