

E-PORTFOLIO

COMPLILATION OF ACADEMIC WORKS

ITC C506-302I

ITE ELECTIVE 3

Final Deliverable

Shan Hiro Rosario



EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

PRELIMS

ACTIVITY

KNN

A Report on the Classification of Iris Species using K-Nearest Neighbors (KNN) Evaluation and Visualization of Model Performance

Shan Hiro Rosario
Department of Computer Science and Engineering
Jose Rizal University
Mandaluyong City, Philippines
shanhiro.rosario@my.jru.edu

Abstract: This study presents the application of the supervised machine learning K-Nearest Neighbor (KNN) technique in assessing and evaluating a classic dataset used for machine learning and statistics. The use of Python libraries Matplotlib and scikit-learn are the core technologies used in the implementation of the KNN algorithm. The dataset has been divided into training and testing the model. KNN, by grouping similar patterns and predicting future outcomes provides valuable insights in how it can be used for predictive modeling by using the Iris dataset that includes 150 samples of iris flowers, with 50 samples from each of the three species (Setosa, Versicolor and Virginica). The study aims to provide an approach for implementing the KNN model, in interpreting and observing its capacity to recognize pertinent patterns, visualizing the results and assessing the classification accuracy.

Keywords—Machine Learning, K-nearest Neighbor, Supervised Learning, Python, MATLAB, Scikit-learn, Predictive Modeling.

I. INTRODUCTION (HEADING I)

Numerous machine learning algorithms have been developed in the last decade and it is also seen in the recent years, specifically during the pandemic in how effective these algorithms are in its predictive qualities and capacities. It is essential in the future events to facilitate the empirical study of these machine learning models as the technology transitions into a new age. The predictive modelling used in this research is the K-Nearest Neighbors (KNN) which, in essence it groups similar patterns in order to predict future outcomes. Combining this with the Iris dataset and the commonly used Python programming language and leveraging its rich capabilities in the application of prediction modelling using real world data. It must be immediately corrected, however, that the potential and restrictions of this method to predictive modelling will require more research as the capacity of the algorithm to recognize pertinent patterns and accuracy in its forecasts changes with differing datasets.

■ KNN ACTIVITY AND PEER REVIEW

Additionally, the K-Nearest Neighbors is often called a

II. REVIEW OF RELATED LITERATURE

According to Rani & Vanishtha, 2017 [1]. Despite KNN's advantages in its simplicity and intuitive implementations of low training time, as well as its predictions based on the proximity of test data points to training data points, it must be understood that it faces challenges in high computational complexity and memory requirements. This is not an indication, however, that KNN has not been utilized and shown promise in solving real-world issues that call for a substantial volume of data. Recent research indicates that the use of KNN in public health, particularly in relation to tuberculosis, it has obtained a final result of 80% accuracy (Delima Sitanggang et al., 2024) [2]. In the hospital environment of patient satisfaction, the KNN technique has seen impressive performance in its capabilities of more than 90% accuracies in various metrics conducted by the researchers Novia Sibutar-Butar et al., 2024 [3].

The KNN model has also seen constant improvement by various researchers in which, its classifiers has been modified for better prediction accuracy despite maintaining its simplicity to implement (Hoque et al., 2021) [4]. The continuous improvement, application and assessment of the K-Nearest Neighbors algorithm, along with its ease of use provides the best foundational opportunity on learning the multidisciplinary field of machine learning. Along with the excellent libraries of the Python programming language and its capabilities in analytics and algorithms, when combined becomes pivotal in gaining thorough understanding and insights in the complexities of Machine Learning.

III. IMPLEMENTATION

The use of Python programming language and its corresponding libraries matplotlib and scikit-learn has been used throughout the testing and training of the model. Along with the Iris dataset imported from the library, in which is a classic dataset in machine learning and statistics. It is commonly used by new practitioners and beginners in the field of computer science, information technology, or

EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

PRELIMS

ACTIVITY

LINEAR REGRESSION

PT-P3 (Implementing Simple and Multiple Linear Regression)

Due Feb 17 by 10:30am Points 0 Submitting a file upload File Types ipynb and docx Available Feb 6 at 12am - Feb 17 at 10:30am

This assignment was locked Feb 17 at 10:30am.
No additional details were added for this assignment.

« Previous

Submission

✓ Submitted!

Feb 15 at 7:14pm

Submission Details

Download PT-P3-

Rosario_ShanHiro.docx

Download

PT_P3_Simple_Multiple_Regressio

Grade: 80 (0 pts possible)

Graded Anonymously: no

Assigned Peer Reviews

✓ Anonymous User

✓ Anonymous User

✓ Anonymous User

Comments:

Hi Rosario, First I hope you are doing well. I am glad to be assigned to your reviewers. I appreciate the opportunity to review your work and learn from your approach to data preprocessing and predictive modeling. Your research provides a structured analysis of data preprocessing techniques and regression modeling using the DMV-CAR Dataset. This peer review will highlight the strengths, weaknesses, and areas for improvement in your study.

One of the most commendable aspects of your study is the well-structured organization of your paper. Upon reviewing your work, I found it easy to

PT-P3 (Implementing Simple and Multiple Linear Regression)

Feb 17 | 0 pts

↳ Required Peer Review 1
Anonymous Student

↳ Required Peer Review 2
Anonymous Student

↳ Required Peer Review 3
Anonymous Student

■ MLR DOCUMENTATION AND PEER REVIEW

EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

PRELIMS

MLR DOCUMENTATION

ACTIVITY

LINEAR REGRESSION

A Report on the Data Preprocessing Pipeline and Predictive Modeling using Simple and Multiple Linear Regression

Shan Hiro Rosario
Department of Computer Science and Engineering
Jose Rizal University
Mandaluyong City, Philippines
shanhiro.rosario@my.jru.edu

Abstract: This study contains an analysis of the Data preprocessing and the best predictive pipeline by conducting a rigorous experiment on the most suitable correlation between the available attributes and its records after the data has been thoroughly cleaned and preprocessed. This assessment is done through python and by using the Large-Scale Dataset for Automotive Applications, DMV-CAR Dataset which is designed for research in relation to car designs such as appearances, consumer analytics and sales modeling.

Keywords—Machine Learning, Linear Regression, Supervised Learning, Python, MATLAB, Scikit-learn, Predictive Modeling.

I. INTRODUCTION

Linear Regression has been used in several and multiple industries concerning quantitative variables such as continuous and discrete values. It however, cannot process Qualitative or categorical variables and the study encoded such variables in binary values of 0 and 1 with the corresponding labels to ensure that no information has been left out and a thorough review, analysis and assessment is performed fairly in the study. A regression model, which is the predictive modeling we are to base the report paper from, is a supervised machine learning model which focuses on the relationships of independent and dependent variables. Also called predictor and predicted variables respectively, as the Machine Learning Model can only accurately predict the labeled data it has been trained to. An extensive application of the regression model has seen in most industries because most businesses' data are linear in nature. [1] Examples include linear forecasting because of its simplicity in the domains of Sales, whereas the analysis focuses on future sales based on advertising expenses, the shifting trends of seasons, and customer footfall among others used in analytics.

[2] Advanced linear models would include Time Series Forecasting models which still uses linear regression as an algorithm but it would be extensively used by researchers that are comparing these regression to others that are deeper and

II. REVIEW OF RELATED LITERATURE

[3] Recent studies have shown that the use for Multiple Linear regression at predicting significant attributes that contributes to Sales on car-related businesses is effective in high level decisions that are made by executives of a company especially for giant businesses that operates in multi-national scale, the ability to forecast, monitor and extract patterns from historical data is a quality standard for any company in the business to ensure that they have their own competitive advantage—to stay ahead of the industry. [4] Furthermore, the impact of such global giants does not only affects the business industry but also the socioeconomic factors of a nation as a research conducted by Daqian L. et al, spatial regression models have been used to analyze several socioeconomic variables in order to gather the income factors, distributions of car sales, and spacial effect across multiple cities in China.

Regression analysis are also effective in financial applications because of its ability to estimate the relationships between the predictor and predicted variable and in this study we would be using its most basic form, as duly explained by Kamil, M. and Kai, S. the principles of covariance and correlation[5] in which complex patterns are then tried to be extracted with more predictor (independent) variables being used to squeeze out the related data and improve the accuracies of the prediction. This method relies on quantitative values as explained in the introductory part as categorical variables are not used in regression modeling.

III. PREPROCESSING

The use of Python programming language and its corresponding libraries matplotlib and scikit-learn has been used throughout the testing and training of the model. Along with the DMV-CAR Dataset that is open-sources and authored by Jingmin Huang, in which its purpose is to support the automotive industry extensively by allowing practitioners of data science to use the dataset for appearance design, consumer analytics, sales forecasting and more.

EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

PRELIMS

- ALL WERE DONE F2F

QUIZ

QUIZ

Set A (II)

47568

JOSÉ RIZAL UNIVERSITY

80 SHAW BOULEVARD
MANDALUYONG CITY

EXAMINATION BOOK

NAME: Shan Hiro T. Rosario

DATE: Q1 - 13 - 2025

INSTRUCTOR: Dr. (Redacted) Paga

SUBJECT: IIE Elective 3

DO NOT TEAR ANY PAGE OUT OF THIS
EXAMINATION BOOK

EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

PRELIMS

- ALL WERE DONE F2F

The image shows a digital quiz interface with a yellow 'QUIZ' button and a dark grey 'QUIZ' button. Below the interface is a photograph of a handwritten document. The document has a header that reads 'IIE Elective Prelim Long Quiz Set A Predictive Modelling'. The main body of the document contains handwritten text discussing data preprocessing and cleaning for a predictive modeling task, specifically mentioning KNN and handling inconsistent values.

IIE Elective Prelim Long Quiz Set A Predictive Modelling

1. In the given dataset, we are presented with customer purchase history. However, the features presented in the attributes and records are using binary numbers of 1 and 0 which is equivalent to "bought" and didn't buy respectively. We must first understand the business model thoroughly before implementing any kinds of predictive modelling practices, which is extremely important before preprocessing the dataset so we could analyze which attributes are to be labelled before feeding it to the KNN algorithm. For example, the dataset records and attributes must have been categorical variables originally, but since the KNN algorithm works best using quantitative variables, it has been encoded to binary values with the corresponding labels. From this assessment we can already derive conclusions that values that are not 0 or 1 must be filtered because it is inconsistent or dirty. To illustrate, Product Category 1 in the row of Customer ID 2 has a value of 100. This is wrong base on the analysis performed earlier. It is inconsistent, and so cleaning techniques must be implemented. The sample dataset only has 5 rows in which it is not realistic to be trained in any model following standard practices. The dataset must have contained tens of thousands of records, whereas directly deleting the inconsistent or missing value should suffice if and only if we can afford it. To illustrate, if we have a dataset of 100k records but due to missing values and we delete it then we are left with 20k records, it becomes unsuitable for predictive modelling. To prevent such scenarios, we can take the average numbers of the column the missing records falls through and

EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

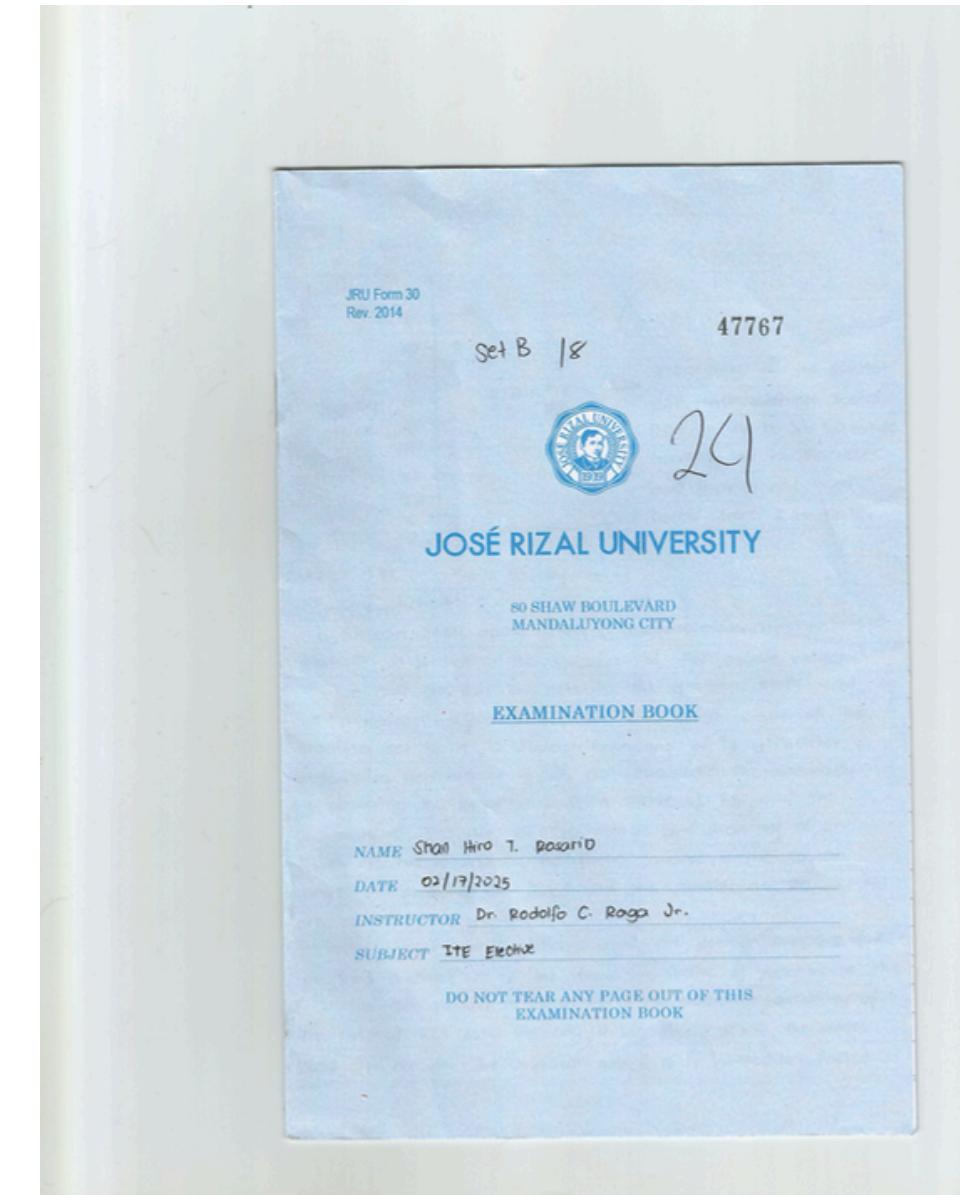
PRELIMS

■ ALL WERE DONE F2F



EXAM

EXAM



EPORTFOLIO

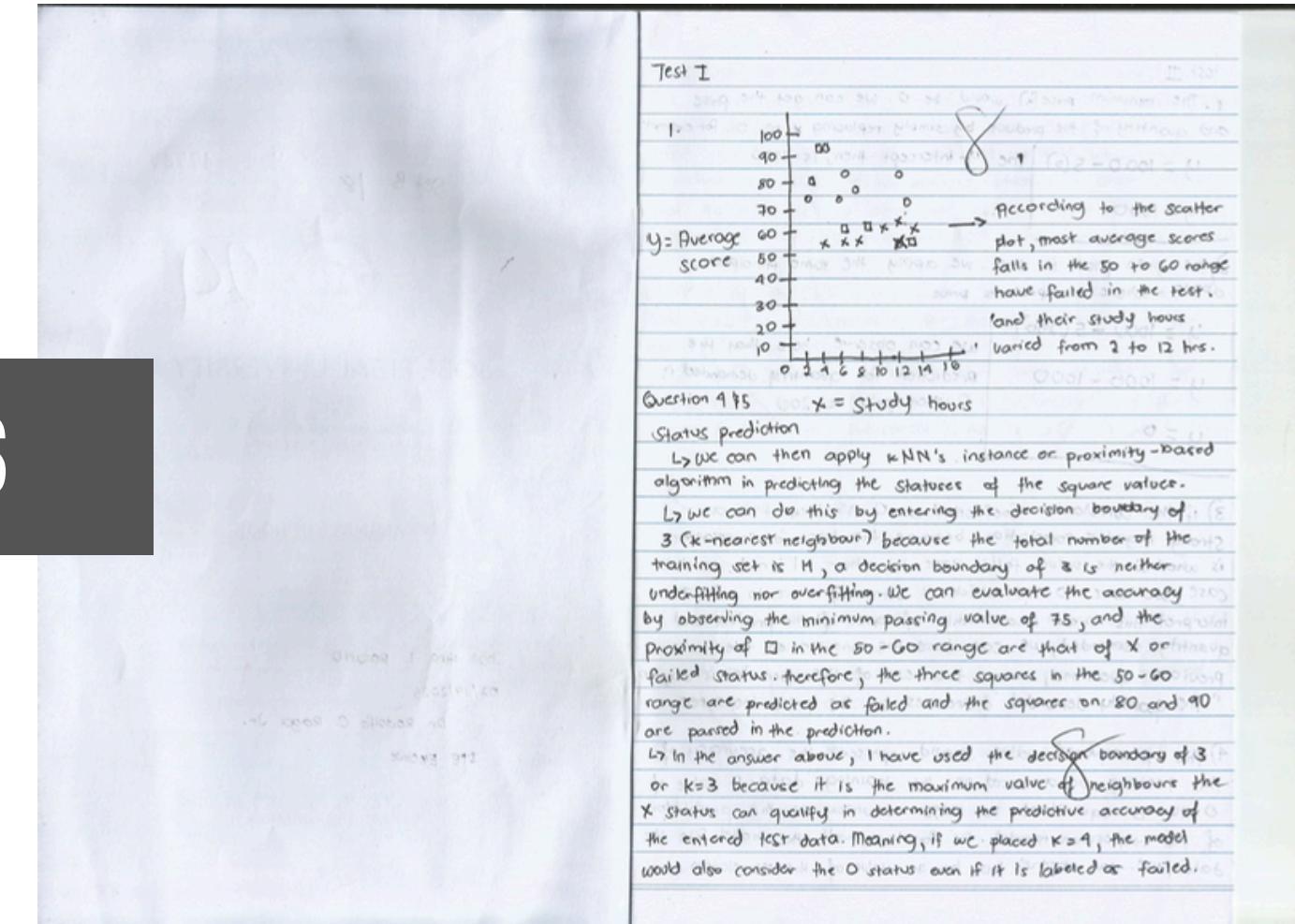


ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

PRELIMS

■ ALL WERE DONE F2F



EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

REFLECTION

MACHINE LEARNING

The introduction of machine learning gave profound insights regarding the future career possibilities of which are data analysts, and the practical implementation of the traditional algorithm of multiple linear regression and kNN realizes the path slowly but surely. The foundations of machine learning lies in the understanding of the algorithms themselves, how they work, how they are implemented through the preprocessing of the dataset and training the model without delving too much on the learning of the python code itself because that is for another field of study, by understanding the concepts and theories, we were able to gain specialized knowledge of the study such as the linear coefficients, the k-fold and train-test split, evaluation metrics for regression tasks, confusion matrices and the outlier removal techniques/methods.

ITE ELECTIVE REFLECTION

EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

MIDTERMS

ACTIVITYDECISION TREES

PT-M1 (Comparing Logistic Regression with Decision Tree for Multinomial Classification)

Due Mar 19 by 11:59pm Points 0 Submitting a file upload File Types ipynb, docx, and xlsx Available Mar 10 at 12am - Mar 19 at 11:59pm

This assignment was locked Mar 19 at 11:59pm. No additional details were added for this assignment.

* Previous Next *

Grade: 85 (0 pts possible)
Graded Anonymously: no
Assigned Peer Reviews
✓ Anonymous User
✓ Anonymous User
✓ Anonymous User

Comments:
The first thing I noticed about the KCE report is its lack of clarity in the abstract. It does not clearly state the problem, methodology, or key findings of the report.
The section on logistic regression requires clarification. The explanation provided is somewhat redundant and unclear. It would be better if only Multinomial Logistic Regression was chosen over Binary Logistic Regression and justify the selection of 'sag' as the solver.
In the preprocessing step, you mentioned using train-test split and K-Fold.

PT-M1 (Comparing Logistic Regression with Decision Tree for Multinomial Classification)
Mar 19 | 0 pts

↳ Required Peer Review 1
Anonymous Student

↳ Required Peer Review 2
Anonymous Student

↳ Required Peer Review 3
Anonymous Student

■ DECISION TREES & LOGISTIC REGRESSION DOCUMENTATION AND PEER REVIEW

EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

MIDTERMS

ACTIVITY

DECISION TREES

Comparative Analysis on Decision Trees and Multinomial Logistic Regression for Sentimental Analysis on a Social Media Platform

Shan Hiro Rosario
Department of Computer Science and Engineering
Jose Rizal University
Mandaluyong, Philippines
shanhiro.rosario@my.jru.edu

Abstract—An observation and step-by-step implementation on the applications of Multinomial Logistics Regression and Decision Trees Classifier in the Sentimental Analysis of Tweets using Python and the twitter dataset from Kaggle.

Keywords—Twitter Sentimental Analysis, Multinomial Logistic Regression, Decision Trees Classifiers, Python, Kaggle Dataset.

I. INTRODUCTION

It must be noted that there are hundreds of millions of users in Twitter, or currently known as X, and so inevitably becomes the center of attention in many businesses and corporations for it is an opportunity in increasing their sales through advertising and analyzing customer behaviors. This paper is created with that in mind, as a real-world application of machine learning in the field of businesses by analyzing an individual's studies, ideas, emotions and their general opinions in a popular social media platform using a supervised machine learning algorithms[1]. We will see be in detail on how we achieve the predictions of the model and its accuracies regarding the sentiments of a tweet, classified through their "Categories" in which is labeled on whether the tweet is in the fields of Video Games, News, Sports and etc. We would provide our evaluation, comparison and our results in the sections further below.

Logistic Regression have been implemented through its multinomial capacities in which we used the Sentiments column as the target attribute on whether it is positive, negative, or neutral and two other independent attributes of Category and Tweets description, therefore turning it into a Multinomial Logistic Regression Model. This is a widely used statistical model because of its feature as easily interpretable because of the Sigmoid Function, rather than its counterpart of linear regression where we interpret the results based on sheer numbers of coefficients making it hard to understand even with the use of graphs. In addition, the main difference between the two is that Logistic Regression's response variable is binomial, which means that the result of predictions are either true or false, correct or wrong[2].

Gini Index

$$Gini = 1 - \sum_{i=1}^n p_i^2$$

Fig. 2

II. DATA COLLECTION

A. Source

■ DECISION TREES & LOGISTIC REGRESSION DOCUMENTATION AND PEER REVIEW

EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

MIDTERMS

ACTIVITY

PROJECT

PT-M2 Research Proposal Writing and Presentation

Due Mar 27 by 9:30am Points 0 Submitting a file upload File Types docx and pptx Available until Mar 27 at 9:30am

This assignment was locked Mar 27 at 9:30am. No additional details were added for this assignment.

◀ Previous Next ▶

Submission
✓ Submitted! Mar 27 at 1:20am Submission Details Download Rosario_Satumba_Yap_PT-M2-ResearchProposal.docx Download RosarioSatumbaYap_PT-M2_ResearchPresentation.pptx

Comments: No Comments

ITE ELECTIVE

Project Proposal

Social Challenges and Opportunities

Research Objectives

Data Collected from Kaggle

Model Development

Model Development

Project Proposal

27 March, 2025

■ PROJECT PROPOSAL

EPORTFOLIO



ELECTIVE PORTFOLIO

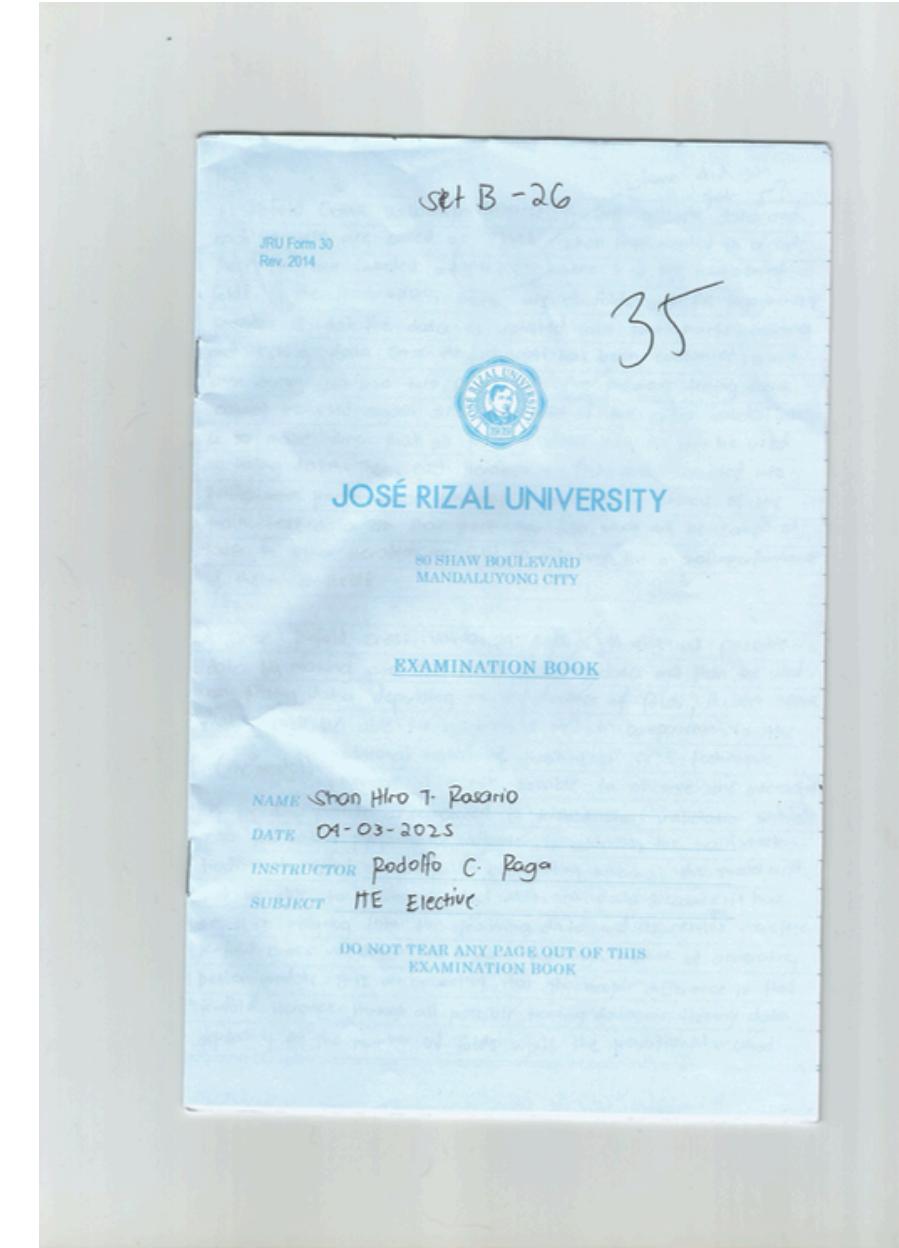
- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

EXAM

DELIVERABLE

MIDTERMS

■ NO QUIZ ON MIDTERMS



EPORTFOLIO

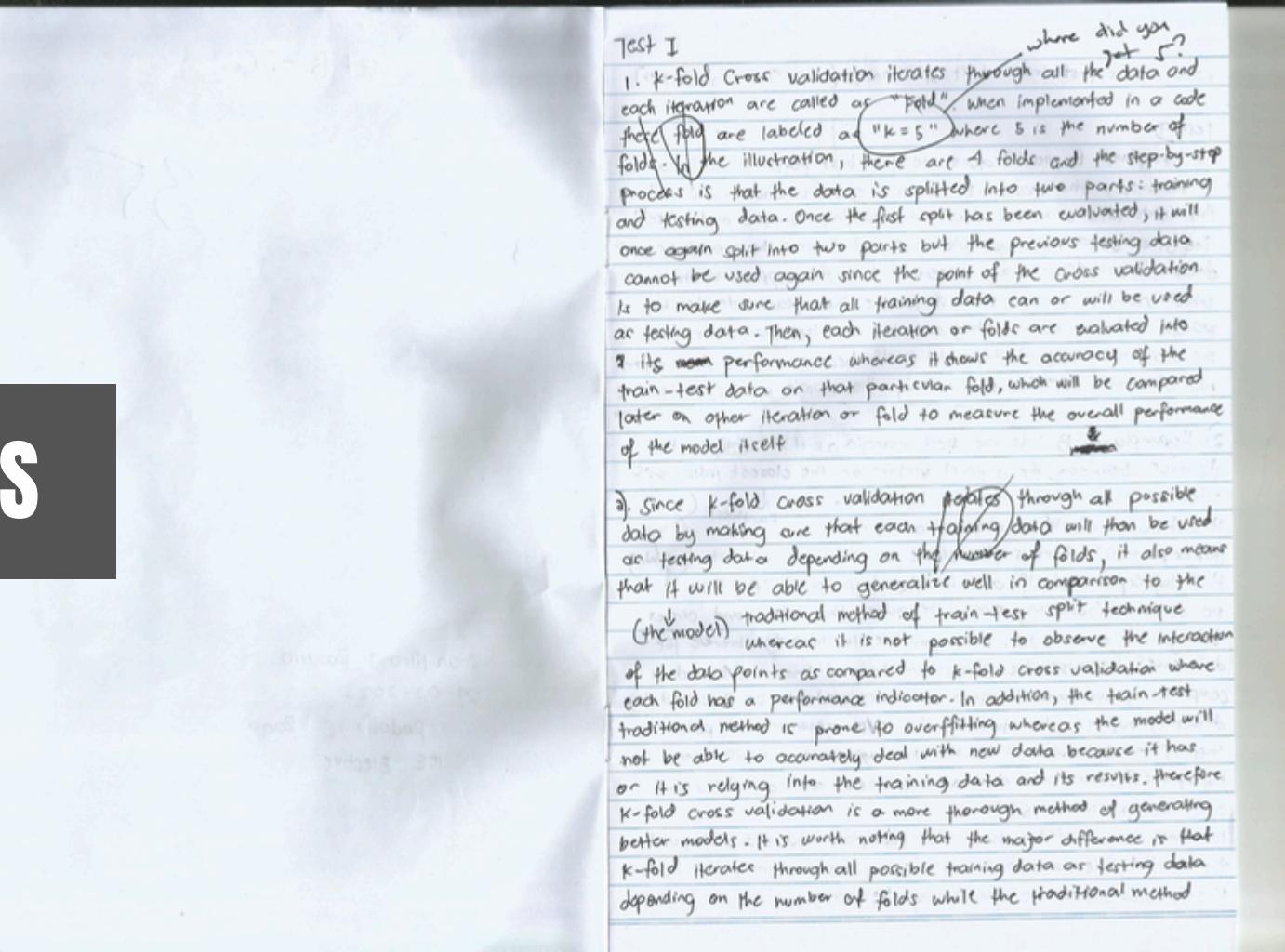


ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

MIDTERMS

■ NO QUIZ ON MIDTERMS



EXAM

DELIVERABLE

EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

REFLECTION

CLASSIFICATION

The midterms mostly had classification tasks by using the algorithms Decision Trees and Logistic regression, along with their respective hyperparameters, preprocessing, confusion matrices, categorical values on the dataset, evaluation metrics such as F1 score, and experimental logs to determine and provide evidence of which model exactly were observed and analyzed to become the best-performing model or fully optimized. A comparison analysis were made on the two algorithms: DT and Logistic regression, while we did not perform any documentations or activities on another algorithm, the Support Vector Machine, we thoroughly studied and reviewed its inner workings such as the vector line, the support vectors that separates the data points in order to see if the chosen vector line is considered as best fit and so on. This is due to the limited time available to us because of the university week and several holidays. Regardless, classification tasks were mostly different than regression tasks that were done in prelims due to its binary or multinomial output that determines the result rather than continuous outcomes.

ITE ELECTIVE REFLECTION

EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

FINALS

ACTIVITY

ENSEMBLE

PT-F1 : (Ensemble Learning for Prediction Modelling)

Due: Apr 21 by 10:30am Points: 0 Submitting a file upload File Types: ipynb, docx, and xlsx Available: Apr 10 at 12am - Apr 21 at 10:30am

This assignment was locked Apr 21 at 10:30am. No additional details were added for this assignment.

Next >

Submission

✓ Submitted! Apr 18 at 10:42pm Submission Details Download PT_F1_(Traditional)_Rosario_ShanHi.ipynb Download PT_F1_BaggingStacking_Yap_Jeo.ipynb Download PT_F1_Boosting_Satumba_Lebron.ipynb Download RosarioShanHirosLogFile-traditional.xlsx Download PT-F1(Ensemble Learning)-Excelsorators.docx

Assigned Peer Reviews

✓ Anonymous User ✓ Anonymous User ✓ Anonymous User

Comments:

The Excelsorators groups aims to predict monthly temperatures in the country's cities using ensemble learning. Their target is to mitigate the heatwaves effect by providing better forecasting tools. The paper is well-written and relevant. It addresses a real-world problem making it relevant and practical. It has a comprehensive Methodology that shows a strong grasp of preprocessing, model selection and evaluation considering every crucial details. But with those strengths there still some are for improvements like limited dataset timeframe the data that was used was solely from the year 2024, but in the

PT-F1 : (Ensemble Learning for Prediction Modelling)
Apr 21 | 0 pts

↳ ⚙ Required Peer Review 1
Anonymous Student

↳ ⚙ Required Peer Review 2
Anonymous Student

↳ ⚙ Required Peer Review 3
Anonymous Student

ENSEMBLE LEARNING DOCUMENTATION AND PEER REVIEW

EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

FINALS

ACTIVITY

ENSEMBLE

Monthly Temperature Prediction in the Philippines using Ensemble Learning: A Regression Approach with Three Baseline Models.

Shan Hiro Rosario
Department of Computer Science and Engineering
Jose Rizal University
Mandaluyong City, Philippines
shanhiro.rosario@my.jru.edu

Lebron Satumba
Department of Computer Science and Engineering
Jose Rizal University
Mandaluyong City, Philippines
lebron.satumba@my.jru.edu

Jeo Christopher Yap
Department of Computer Science and Engineering
Jose Rizal University
Mandaluyong City, Philippines
jeochristopher.yap@my.jru.edu

Abstract—Heatwaves has seen a rising trend over recent years. While problems and causes may vary, a necessary solution and mitigations should be implemented to reduce the risks associated with extreme heat. Such causes include but are not limited to: Health, Financial and Mental Stress which inevitably leads to lower productivity and overall efficiency of the human body as sudden rises and fall in temperature lowers our endurance mentally and physically. Through the implementation of a monthly-temperature predictive model using Ensemble Learning, the proponent of the study hopes to offer long-term planning in combating extreme heat waves in the country of the Philippines.

Keywords—Temperature Forecasting, Ensemble Learning, Quantitative Regression, Decision Trees Regression, Multiple Linear Regression, K-Nearest Neighbors, Bagging, Boosting, Stacking.

I. INTRODUCTION (HEADING I)

Extreme heat waves are being experienced throughout the world and have been more observed in recent years. These spikes in temperature and sentiments of the individuals could be monitored through reliable articles[1][2] and being discussed in the academia by various researchers. [3] the most pressing matter lies in the unpredictable mother nature in which countless weather researchers have been constantly monitoring these scenarios however, the recent temperatures have gone well outside of expectations and it specifically impacted the already high temperatures of tropical countries. The Philippines—included to those of negatively affected countries have suffered severe repercussions and the majority of the population bore the brunt of it as a developing country, most individuals could not rely on the conveniences of technology to endure the sudden spikes of temperature as this could lead to financial problems and their health would deteriorate by then.

II. RESEARCH OBJECTIVES

The foremost goal of the proponents of the study is to offer long-term planning of individuals in the philippines by predicting the next 12 Month's of temperature of the chosen city among all the cities in the philippines. The second

performed in order to see the limits of the performance of traditional models, after providing a comparison analysis of each other.

III. METHODOLOGY

The Dataset used in the study are provided by the author Bwandowando in the website Kaggle [4] in which the title are derived as: Philippine Major Cities Weather Data 2024 providing a 3-hour interval of each variables affecting temperature with the used attributes overview are as follows:

main.feels like	-	the temperature felt after considering humidity and wind.
main.temp_min	-	Minimum recorded temperature.
main.temp_max	-	Maximum recorded temperature.
main.humidity	-	Atmospheric moisture (%)
wind.speed	-	Wind velocity (m/s).
clouds.all	-	Cloud coverage (%)
datetime.hour	-	Hour of the day (0-23).
datetime.day	-	Day of the month (1-31).
coord.lat	-	Latitude of the city.
coord.lon	-	Longitude of the city.
datetime.year	-	Year of the data record.

It must be noted that these attributes are used after it has been preprocessed and the selection will be discussed further below. However, the limitations of the dataset was that although, the records spans over 400 thousand, the year it was obtained was limited to the year 2024 and it did not capture the years below them. These data are collected and organized from January to December of Year 2024.

Preprocessing

Data preprocessing are the same throughout all the three models to ensure consistency of the results and experiments. Step-by-step implementation are as follows:

First we imported the raw dataset in the google colab python terminal:

```
from google.colab import files
uploaded = files.upload()

# Check uploaded files
uploaded
```

ENSEMBLE LEARNING DOCUMENTATION AND PEER REVIEW

EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

FINALS

ACTIVITY

MLP

PT-F2 : MLP Challenge

Due May 2 by 11:30pm Points 0 Submitting a file upload File Types docx, ipynb, xlsx, and csv Available Apr 24 at 12am - May 2 at 11:30pm

This assignment was locked May 2 at 11:30pm.
No additional details were added for this assignment.

Previous Next

Submission

✓ Submitted! Apr 30 at 4:04pm Submission Details Download RosarioShanHiro-IEEE-1.docx Download RosarioShanHiro_MLP.ipynb Download RosarioShanHiro-LogFile-MLP.xlsx Download raw_2025_weather_data.csv Download raw_2024_weatherdata.csv

Assigned Peer Reviews

✓Anonymous User ✓Anonymous User ✓Anonymous User

Comments:

In reviewing this paper on "A Regression Approach on Predicting the Monthly Temperature in the Philippines using Multi-Layer Perceptron," Mr. Rosario did a good job of making an organized IEEE paper. He first started by defining a relevant and impactful problem, namely that in the Philippines and increasingly in all other islands of the world, temperature forecasting is a crucial area, to enable accurate temperature forecasting. I find it commendable that the author chose to adopt a data-driven method by using historical temperature data from PAGASA to further authenticate and enhance the study.

PT-F2 : MLP Challenge May 2 | 0 pts

↳ Required Peer Review 1 Anonymous Student

↳ Required Peer Review 2 Anonymous Student

↳ Required Peer Review 3 Anonymous Student

MULTI LAYER PERCEPTRON DOCUMENTATION AND PEER REVIEW

EPORTFOLIO

ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION



FINALS

ACTIVITY

MLP

A Regression Approach on predicting the Monthly Temperature in the Philippines using Multi Layer Perceptron.

Shan Hiro Rosario
Department of Computer Science and
Engineering
Jose Rizal University
Mandaluyong City, Philippines
shanhiro.rosario@my.jru.edu

Abstract—Heatwaves has seen a rising trend over recent years. While problems and causes may vary, a necessary solution and mitigations should be implemented to reduce the risks associated with extreme heat. Such causes include but are not limited to: Health, Financial and Mental Stress which inevitably leads to lower productivity and overall efficiency of the human body as sudden rises and fall in temperature lowers our endurance mentally and physically. Through the implementation of a monthly-temperature predictive model using Neural Network specifically Multi-Layer Perceptron, the proponent of the study hopes to offer long-term planning in combating extreme heat waves in the country of the Philippines.

III. METHODOLOGY

The Dataset used in the study are provided by the author Bwandawando in the website Kaggle [4] in which the title are derived as: Philippine Major Cities Weather Data 2024 providing a 3-hour interval of each variables affecting temperature with the used attributes overview are as follows:

main.feels like – the temperature felt after considering humidity and wind.
main.temp_min – Minimum recorded temperature.
main.temp_max – Maximum recorded temperature.
main.humidity – Atmospheric moisture (%).
wind.speed – Wind velocity (m/s).
clouds.all – Cloud coverage (%).
datetime.hour – Hour of the day (0-23).
datetime.day – Day of the month (1-31).
coord.lat – Latitude of the city.
coord.lon – Longitude of the city.
datetime.year – Year of the data record.

It must be noted that these attributes are used after it has been preprocessed and the selection will be discussed further below. However, the limitations of the dataset was that although, the records spans over 400 thousand, the year it was obtained was limited to the year 2024 and it did not capture the years below them. These data are collected and organized from January to December of Year 2024.

II. RESEARCH OBJECTIVES

The foremost goal of the proponents of the study is to offer long-term planning of individuals in the philippines by predicting the next 12 Month's of temperature of the chosen city among all the cities in the philippines. The second objective is to offer insights to future weather and climate

MULTI LAYER PERCEPTRON DOCUMENTATION AND PEER REVIEW

EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

FINALS

ACTIVITYREPORT

Final Project Status Report with Evidence

Due May 8 by 10:30am Points 0 Submitting a file upload File Types ipynb, docx, xlsx, and pptx Available May 7 at 12am - May 8 at 10:30am

This assignment was locked May 8 at 10:30am. No additional details were added for this assignment.

[« Previous](#)



Model Development

1 Model Development

2 Model Development

3 Model Development

4 Model Development

5 Model Development

EXCELERATORS

Project Progress Report



Submission

✓ Submitted! May 7 at 9:26pm

Submission Details

Download ITC ELECTIVE PROJECT REPORT - EXCELERATORS.pptx

Download

Excellerators_FinalsProject.docx

Download

Excellerators_MLR_notebook.ipynb

Download

Excellerators_RandomForest_notebook.ipynb

Download

Excellerators_MLP_notebook.ipynb

Download

Excellerators_MLPfinal_logs.xlsx

Download

Excellerators_RandomForest_logs.xlsx

Comments:

No Comments

■ FINALS ACTIVITY

EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

FINALS

■ FINALS PROJECT

The screenshot shows a learning management system interface. At the top, there are two tabs: 'PROJECT' (highlighted in yellow) and 'PROJECT' (highlighted in grey). Below the tabs, the page title is 'ITC C506-302I > Assignments > Final Project with IEEE paper'. On the left, a sidebar lists course navigation links: Home, Zoom, Syllabus, Modules, Announcements, Outcomes, Discussions, Collaborations, Chat, Office 365, Google Drive, and Microsoft Teams. The main content area is titled 'Final Project with IEEE paper' and includes details: Due May 16 by 11:59pm, Points 0, Submitting a file upload, Available May 14 at 12am - May 16 at 11:59pm. It also notes that the assignment was locked on May 16 at 11:59pm. A submission section shows a green checkmark next to 'Submitted!' with the date May 16 at 6:51am. A download link for 'Excellers, FinalsProject-1.docx' is available. The right side of the screen shows a vertical list of files for download, including 'Excellers, MLR_notebook-1.ipynb', 'Excellers, RandomForest_notebook-1.ipynb', 'Excellers, MLP_notebook-1.ipynb', 'Excellers, MLP_logs.xlsx', and 'Excellers, RandomForest_logs-1.xlsx'. At the bottom right, there is a 'Comments' section with a note 'No Comments'.

EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

FINALS

■ FINALS PROJECT

PROJECT

PROJECT

Long-term Air Temperature predictions in the Philippines using Supervised Machine Learning Algorithms: A Comparison Analysis Approach.

Shan Hiro Rosario
Department of Computer Science and
Engineering
Jose Rizal University
Mandaluyong City, Philippines
shanhiro.rosario@my.jru.edu

Lebron Satumba
Department of Computer Science and
Engineering
Jose Rizal University
Mandaluyong City, Philippines
lebron.satumba@my.jru.edu

Jeo Christoper Yap
Department of Computer Science and
Engineering
Jose Rizal University
Mandaluyong City, Philippines
jeochristopher.yap@my.jru.edu

Abstract—A Monthly-basis prediction on outdoors temperature using Random Forest Regression, Multiple Linear Regression and Philippine Major Cities Weather Data from 2023 to 2025 of March containing hundreds of thousands of records and 27 attributes. Implementation will be done through Python, Google Colab's Jupyter Notebook, and analysis of the societal challenges, in which, the model would be developed for individuals in order to aid them in financial planning such as electric bills, item purchases and awareness of extreme heat that may arise diseases in addition to the increasingly rising temperatures and heatwaves throughout the world.

Keywords—Temperature Forecasting, Multiple linear Regression, Random Forest Regression, Multiple Linear Regression Python, Philippine Weather Dataset.

I. INTRODUCTION

Problem Statement (Societal Challenge & Domain)

Heatwaves are currently being experienced throughout the world[1] and the tropical countries are severely affected due to the concerning levels of temperature that leads to shut downs of educational institutions[2] in order to avoid health risks such as dehydrations, stroke, heat exhaustion, and various skin diseases in addition to greater financial strain on electrical costs due to the reliance on air conditioning, or fans. This is a risk factor on all individuals, specifically the discomfort on public spaces, such as inconvenience on public transport, waiting areas, workplaces or areas without proper cooling become unbearable in addition to severe health problems and food spoilage. Furthermore, the decrease in productivity due to stress in elevated environmental temperatures goes beyond the inconveniences on an individual and eventually leads into a deterioration in their well-being including but not limited to physical health and mental health[3].

Research Objectives

The proponents strives to provide insights for long-term decision making of individuals and emergency planning based on the temperature predictions of the model developed as the inconveniences and elevated risks of temperature increases have been observed on a positive trend in recent years[4]. Due to this, it is clear that the preparations are needed to be measured in a monthly, or yearly basis in which is the foundation of the research paper.

The model will be based on the 138 major cities that have been recorded in the dataset provided which contains historical data from 2023 to 2025 of march. The beneficiaries would be able to analyze whether the city of their choices is suitable for their own needs such as tourists who plans to visit popular cities that might be a major inconvenience instead due to the high temperatures that would be predicted by the model.

In addition, the beneficiaries would also be able to realize their financial capabilities on electric bills due to the need of air conditioning, electric fans, food and item purchases such as sunscreens and more should they be aware of the high temperature changes that spans over the whole year, on a monthly basis.

II. REVIEW OF RELATED LITERATURE

Monthly temperature and rainfall forecasts have been conducted in Perlis, by Sin Lim, et al. by conducting an analysis of forecasting method such as Simple Seasonal Exponential Smoothing, Holt Winter Additive, Holt Winter Multiplicative and Seasonal ARIMA[5]. Their research has proven that the Simple Seasonal Exponential Smoothing have outperformed other machine learning methods by implementing error metrics such as Mean Absolute Error, Root Mean Square Error and Mean Absolute Percentage Error that is highly relevant to the current research in which it would be observed in the implementation of the predictive model whether it is a suitable methods in determining the error performance of the model.

A research on monthly temperature variations have also

EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
- PERFORMANCE TASKS
- FINAL EXAMINATION
- REFLECTION

FINALS

■ FINALS PROJECT

Research Objectives

A research on monthly temperature variations have also been conducted by Sin Lim, et al. by conducting an analysis of forecasting method such as Simple Seasonal Exponential Smoothing, Holt Winter Additive, Holt Winter Multiplicative and Seasonal ARIMA[5]. Their research has proven that the Simple Seasonal Exponential Smoothing have outperformed other machine learning methods by implementing error metrics such as Mean Absolute Error, Root Mean Square Error and Mean Absolute Percentage Error that is highly relevant to the current research in which it would be observed in the implementation of the predictive model whether it is a suitable methods in determining the error performance of the model.

Abstract—A Monthly-basis prediction on outdoors temperature using Random Forest Regression, Multiple Linear Regression and Philippine Major Cities Weather Data from 2023 to 2025 of March containing hundreds of thousands of records and 27 attributes. Implementation will be done through Python, Google Colab's Jupyter Notebook, and analysis of the societal challenges, in which, the model would be developed for individuals in order to aid them in financial planning such as electric bills, item purchases and awareness of extreme heat that may arise diseases in addition to the increasingly rising temperatures and heatwaves throughout the world.

Keywords—Temperature Forecasting, Multiple linear Regression, Random Forest Regression, Multiple Linear Regression Python, Philippine Weather Dataset.

I. INTRODUCTION

Problem Statement (Societal Challenge & Domain)

Heatwaves are currently being experienced throughout the world[1] and the tropical countries are severely affected due to the concerning levels of temperature that leads to shut downs of educational institutions[2] in order to avoid health risks such as dehydrations, stroke, heat exhaustion, and various skin diseases in addition to greater financial strain on electrical costs due to the reliance on air conditioning, or fans. This is a risk factor on all individuals, specifically the discomfort on public spaces, such as inconvenience on public transport, waiting areas, workplaces or areas without proper cooling become unbearable in addition to severe health problems and food spoilage. Furthermore, the decrease in productivity due to stress in elevated environmental temperatures goes beyond the inconveniences on an individual and eventually leads into a deterioration in their well-being including but not limited to physical health and mental health[3].

II. REVIEW OF RELATED LITERATURE

Monthly temperature and rainfall forecasts have been conducted in Perlis, by Sin Lim, et al. by conducting an analysis of forecasting method such as Simple Seasonal Exponential Smoothing, Holt Winter Additive, Holt Winter Multiplicative and Seasonal ARIMA[5]. Their research has proven that the Simple Seasonal Exponential Smoothing have outperformed other machine learning methods by implementing error metrics such as Mean Absolute Error, Root Mean Square Error and Mean Absolute Percentage Error that is highly relevant to the current research in which it would be observed in the implementation of the predictive model whether it is a suitable methods in determining the error performance of the model.

Long-term Air Temperature predictions in the Philippines using Supervised Machine Learning Algorithms: A Comparison Analysis Approach.

Shan Hiro Rosario
Department of Computer Science and
Engineering
Jose Rizal University
Mandaluyong City, Philippines
shanhiro.rosario@my.jru.edu

Lebron Satumba
Department of Computer Science and
Engineering
Jose Rizal University
Mandaluyong City, Philippines
lebron.satumba@my.jru.edu

Jeo Christoper Yap
Department of Computer Science and
Engineering
Jose Rizal University
Mandaluyong City, Philippines
jeochristopher.yap@my.jru.edu

PROJECT

PROJECT

EPORTFOLIO



ELECTIVE PORTFOLIO

- WRITTEN WORKS
 - PERFORMANCE TASKS
 - FINAL EXAMINATION
 - REFLECTION
-

REFLECTION

MACHINE LEARNING

The final term activities were the implementation of the considerably advanced supervised machine learning algorithms such as the ensemble learning in which it combines several algorithms as the baseline models and it will be trained using ensemble methods such as bagging for overfitting and boosting for underfitting, assigning weighted values to them for further observations and analysis. In addition, neural networks were also explored, specifically the Multi layer perceptron, which consists of weights, hidden layers, output layer, activation functions, and a general feedforward network structure in addition to the number of epochs, learning rate and so on. This neural network uses both regression and classification in its neurons to determine the linear and non-linearity relationships of the dataset it was trained on. Although the algorithm is powerful it must be noted that the computer will be stressed, and so the training or execution time is noticeably slower than other algorithms. There was also a status report for the final project in the finals.

■ FINALS EXAMINATION

Reflection

Throughout the whole semester, i learned the intricacies of machine learning, such as the preprocessing steps, outlier detection, knowledge of supervised learning models, evaluation metrics/techniques, and the experimental logs to observe the best performing model. I believe this is another career path that opened once i graduated, as even with the absence of a computer science degree, the knowledge of applied machine learning would still be relevant in the industry. In addition, the types of exam done through essays or challenging our understanding and logic rather than simulating the traditional multiple choices is a rather unconventional approach that generated benefits as i thoroughly understood and remembered all the necessary foundations needed for supervised machine learning models.

Shan Hiro Rosario

