

Meeting Minutes

Team name: 888 – Peaky Blinders

Meeting No: 4

Location: Meeting room 301, Library, Swinburne University of Technology, Hawthorn, VIC 3122

Attending: Xuan Tuan Minh Nguyen, Trong Dat Hoang, Henry Nguyen

Apologies: None.

Meeting start time: 9:00AM – 26/08/2024.

Matters arising from Previous minutes: Yes No **X**

Issues raised from previous minutes	Discussions	Outcome (Resolved?)

Confirmation of minutes from last meeting: Yes **X** No

Outcome of meeting:

Issue	Discussion in brief	Outcome	Action: Name and Timeline
Acquiring Datasets	<p>In short, we discussed to determine which sources should be a truthy source to acquire the datasets regarding civil aviation topics. And how to ensure that the acquired datasets are “healthy”?</p> <p>Sources: Should the datasets be acquired from open-source platforms like Kaggle, dataset.com, IBM, etc. or should it be from the governments, such as: data.gov.au, casa.gov.au, bitre.gov.au, etc.</p>	<p>All members in the group agreed that datasets should be taken from the government platform since this is the most truthy and has all required information.</p> <p>For the healthiness of the datasets, the group decided to go with flight prices as the metrics to evaluate the healthiness of the datasets.</p>	Xuan Tuan Minh Nguyen: Responsible for collecting the datasets from the government source and evaluating the “healthiness” of the datasets by Wednesday (August 28th, 2024).

	<p>Healthiness: In which case do we determine that the datasets are fit with the given context?</p> <p>Is it by flight price, or by flight delay?</p>		
Handle Missing Data	<p>In short, we discussed on how the missing parameters of the dataset could be resolved and how do we deal with it without breaking the datasets.</p> <p>What field of datasets are actually missing the data? Is it all right if we fully drop the columns or perform changes to the columns but does not affect the dataset?</p> <p>If not, how could we handle with missing values in the datasets?</p>	<p>All members agree that we should evaluate if the column is appropriate to perform changes, if there are lots of missing data in lots of columns, we could consider dropping that dataset and move to the other one.</p> <p>Using some techniques, such as: Imputation data, handling missing data without alteration, etc. to effectively handle missing data.</p>	Henry Nguyen: Responsible for handling the missing datasets and decide to drop the datasets if there are lots of missing data by Friday (August 30th, 2024).
Remove unnecessary duplications	<p>In short, we discussed to find out the solution to remove some duplications that are unnecessary.</p> <p>How do we find out if there are duplications and determine if the that duplication is essential to the datasets or not?</p> <p>What are the solutions to drop the unnecessary duplications from the datasets and how do we automate that ?</p>	Although all members agreed that the dataset should not have any duplications and our group should have an automated pipeline to clean the unnecessary duplications, we have gone to the decision of not making the automated pipeline at the moment since it is time consuming and not essential for the current assignment.	Henry Nguyen: Responsible for cleaning the unnecessary duplications in the datasets by Friday (August 30th, 2024).
Correct Inconsistences	In this section, we discuss on what should we deal with the data	Although all members agreed that the inconsistencies inside	Trong Dat Hoang: Responsible for double checking and fixing the

	<p>that is not correct, or in specific, hard to work with programming and training data.</p> <p>How do we convert the dataset into the field that it should be easy to work with programming and training the machine learning model? Such as resolving if the flight hours are 20+ hours but the date are count as 2 days (Could be truthy or falsy).</p> <p>What are the solutions to automate this process ?</p>	<p>the datasets should be converted into a machine-friendly format, automated pipeline to perform this task is not eligible at the moment since it is time consuming and not crucial for the current assignment.</p>	<p>inconsistences by Sunday (September 1th, 2024).</p>
--	---	--	--

Actions in brief:

- Xuan Tuan Minh Nguyen: Responsible for collecting the datasets from the government source and evaluating the “healthiness” of the datasets by Wednesday (August 28th, 2024).
- Henry Nguyen: Responsible for handling the missing datasets, deciding to drop the datasets if there are lots of missing data, and cleaning the unnecessary duplications in the dataset by Friday (August 30th, 2024).
- Trong Dat Hoang: Responsible for double-checking and fixing the inconsistencies by Sunday (August 1st, 2024).

Meeting closed at: 10:45AM – 26/08/2024.

Next meeting time, date and location: 9:00AM – 02/09/2024. Meeting room 301, Library, Swinburne University of Technology, Hawthorn, VIC 3122.

Meeting Minutes

Team name: 888 – Peaky Blinders

Meeting No: 5

Location: Meeting room 301, Library, Swinburne University of Technology, Hawthorn, VIC 3122

Attending: Henry Nguyen, Xuan Tuan Minh Nguyen, Trong Dat Hoang

Apologies: None

Meeting start time: 9:00AM – 02/09/2024.

Matters arising from Previous minutes:

Yes X

No

Issues raised from previous minutes	Discussions	Outcome (Resolved?)
The found datasets have too many missing data	There are lots of missing data that occurred on the given datasets, such as there are no connecting information for flight that has more than one stop and more than 20+ hours of flight time which mean inconsistent for cleaning and training.	Move on to another government-based datasets, which is cleaner and easier to apply to the current context. Resolved.
The datasets are not consistent for training.	Since the found datasets are inconsistent, the corrected inconsistencies seem to be unreliable.	The new datasets are easier to clean and correct the inconsistencies. Resolved.

Confirmation of minutes from last meeting:

Yes X

No

Outcome of meeting:

Issue	Discussion in brief	Outcome	Action: Name and Timeline
Standardize the data and resolve conflictions	Discuss how to standardize the data and what solutions should be made. For example, for the date columns, how do we converted from Sep 1th 2024 to 2024-09-01 format, so that it is easier to let model interact with the datasets without conflictions	For the standardization, create some helper functions to check if the data is not in the right format, if not then converted it into the right format.	Xuan Tuan Minh Nguyen: Responsible for evaluate and write the helper functions to standardize the data by Tuesday(September 3rd, 2024).

Normalize features	<p>Discuss what are the suitable ways and which tools (libraries, frameworks, etc) that we could use to normalize the features.</p> <p>How do we effectively normalize all the features, especially numerical features to 0 and 1 so that the model could learn and predict effectively?</p>	For the normalization, use the MinMaxScaler from sklearn to normalize the features into the number of 0 and 1	Trong Dat Hoang: Responsible for evaluate and write the helper functions to normalize the data by Wednesday (September 4th, 2024).
Split categorized data into train and test data for models to train and evaluate	Discuss which tools to use for splitting and categorizing data into train and test data. And which metrics do we used for splitting data?	Use train_test_split function from sklearn to split 80% of data for training and 20% of data for testing. Split both for classification and regression.	Henry Nguyen: Responsible for evaluate and write the function to split the data by Friday (September 6th, 2024).
Create automated process for fetching and cleaning data	This category is being moved into the Assignment 3	This category is being moved into the Assignment 3	This category is being moved into the Assignment 3

Actions in brief:

- Xuan Tuan Minh Nguyen: Responsible for evaluate and write the helper functions to standardize the data by Tuesday(September 3rd, 2024).
- Trong Dat Hoang: Responsible for evaluate and write the helper functions to normalize the data by Wednesday (September 4th, 2024).
- Henry Nguyen: Responsible for evaluate and write the function to split the data by Friday (September 6th, 2024).

Meeting closed at: 10:30AM – 02/09/2024.

Next meeting time, date and location: 9:00AM – 09/09/2024. Meeting room 301, Library, Swinburne University of Technology, Hawthorn, VIC 3122.

Meeting Minutes

Team name: 888 – Peaky Blinders

Meeting No: 6

Location: Meeting room 301, Library, Swinburne University of Technology, Hawthorn, VIC 3122

Attending: Henry Nguyen, Xuan Tuan Minh Nguyen, Trong Dat Hoang

Apologies: None

Meeting start time: 9:00AM – 09/09/2024.

Matters arising from Previous minutes:

Yes X

No

Issues raised from previous minutes	Discussions	Outcome (Resolved?)
Data standardization should be performed, and conflicts need to be resolved.	The functions that help standardize date formats and categorical variables were examined.	Successfully resolved all standardized data format. Resolved
Categorized the data and then split it into train and test datasets.	The data splitting function for creating training and testing sets has been finished.	Data split for training and for testing. Resolved.

Confirmation of minutes from last meeting:

Yes X

No

Outcome of meeting:

Issue	Discussion in brief	Outcome	Action: Name and Timeline
Categorize Variables	We discussed about the effective categorization of variables and intend to utilize different types of variables.	Agreed a plan for classifying variables using particular encoding techniques.	Henry Nguyen: Responsible for implementing the categorization by Tuesday (12/09/2024).
Create Stream and Data Pipeline	Discussed about creating and putting into action an automated	Pipeline structure has been initially planned, and the	Trong Dat Hoang: Develop and test the pipeline by Friday (15/09/2024).

	data pipeline for streaming and pre-processing data.	implementation is scheduled.	
Split train/test Data	Find out the appropriate tools and metrics to use when dividing data into training and testing sets for later model assessment.	The function for splitting the data into training and testing sets has already been developed, and there are additional enhancements in the pipeline.	Xuan Tuan Minh Nguyen: Review and optimize data splitting by Wednesday (13/09/2024).
Identify Features/Libraries	Explored possible features and libraries to use for advanced data analysis and feature engineering.	Core libraries have been identified, but additional analysis is needed.	All team members: Research additional Features and libraries by Week 7 meeting.

Actions in brief:

- Henry Nguyen needs to have the categorization methods using encoders implemented by Tuesday (12/09/2024).
- Xuan Tuan Minh Nguyen is responsible for optimizing and reviewing the data splitting process by Wednesday (13/09/2024).
- Trong Dat Hoang has to develop and test the automated data pipeline by Friday (15/09/2024).
- All team members are expected to conduct research and propose new features and libraries before the next meeting.

Meeting closed at: 11:30AM – 09/09/2024.

Next meeting time, date and location: 9:00AM – 16/09/2024. Meeting room 301, Library, Swinburne University of Technology, Hawthorn, VIC 3122

Meeting Minutes

Team name: 888 – Peaky Blinders

Meeting No: 7

Location: Meeting room 301, Library, Swinburne University of Technology, Hawthorn, VIC 3122

Attending: Henry Nguyen, Xuan Tuan Minh Nguyen, Trong Dat Hoang

Apologies: None

Meeting start time: 9:00AM – 06/09/2024.

Matters arising from Previous minutes:

Yes X

No

Issues raised from previous minutes	Discussions	Outcome (Resolved?)
Categorized variables.	Categorize variables, and implementation successfully ensures that the model receives properly formatted data.	Variables successfully categorized. Resolved.
Split train and test Data	The data splitting function has been optimized to achieved a balanced distribution between the training and testing datasets.	Data splitting reviewed and improved. Resolved.
Create Stream and Data Pipeline	Additional improvements are necessary to manage intricate data situations.	The pipeline has been tested, but it needs further refinement. Some progress has been made in resolving the issue. Not Fully Resolved.

Confirmation of minutes from last meeting:

Yes X

No

Outcome of meeting:

Issue	Discussion in brief	Outcome	Action: Name and Timeline
Identify Features/Libraries	Explored different functionalities and frameworks to improve the model's abilities. Utilize techniques for selecting features and transforming data.	Chosen important characteristics and software packages for in-depth analysis.	Henry Nguyen: Implement feature selection methods by Tuesday (19/09/2024).

Research Model	Explored different machine learning models that are appropriate for the dataset, with an emphasis on regression and classification.	Team have found some potential models to evaluate in the upcoming phase.	Trong Dat Hoang: Finalize model research by Thursday (21/09/2024).
Select Machine Learning Model	Went over the factors to consider when choosing the ultimate machine learning model, which encompassed performance measurements.	Selection process underway; criteria established.	All team members: Evaluate models and finalize selection by Friday (22/09/2024).
Train Models	Scheduled the training sessions and allocated resources for training different models using the dataset.	The training will commence as planned next week.	Henry Nguyen: Prepare data for training by Friday (22/09/2024).

Actions in brief:

- The training will commence as scheduled. Henry Nguyen is tasked with implementing feature selection methods by Tuesday (19/09/2024) and preparing data for training by Friday (22/09/2024).
- Xuan Tuan Minh Nguyen is responsible for developing and testing new features by Wednesday (20/09/2024).
- Trong Dat Hoang is expected to finalise research on machine learning models by Thursday (21/09/2024).
- All team members must evaluate and finalise the machine learning model selection by Friday (22/09/2024).

Meeting closed at: 11:00AM – 16/09/2024

Next meeting time, date and location: 9:00AM – 23/09/2024. Meeting room 301, Library, Swinburne University of Technology, Hawthorn, VIC 3122

Meeting Minutes

Team name: 888 – Peaky Blinders

Meeting No: 8

Location: Meeting room 301, Library, Swinburne University of Technology, Hawthorn, VIC 3122

Attending: Xuan Tuan Minh Nguyen, Trong Dat Hoang, Henry Nguyen

Apologies: None

Meeting start time: 9:00AM – 23/09/2024.

Matters arising from Previous minutes:

Yes X

No

Issues raised from previous minutes	Discussions	Outcome (Resolved?)
Research model	Team have completed the selection of potential machine learning models for both classification and regression tasks and also created a comparison matrix that is based on performance metrics.	Ready for model selection. Resolved.
Identify features/library	Applied feature selection techniques. Team have assessed different libraries for preprocessing data and engineering features.	Primary libraries have been chosen and essential features have been identified. Resolved.

Confirmation of minutes from last meeting:

Yes X

No

Outcome of meeting:

Issue	Discussion in brief	Outcome	Action: Name and Timeline
Select Machine Learning Model	The comparison matrix has been examined for potential models and decided to begin initial testing with three different models.	The models chosen for testing and assessment.	Trong Dat Hoang plans to start testing models using the prepared data by Wednesday (27/09 2024).
Train model and tune parameters	Talked about the training plan and the resources needed. Team agreed to use iterative training	The tuning process is scheduled to implement after the	Xuan Tuan Minh Nguyen needs to have the tuning scripts and methodologies prepared. By Friday (29/09/2024).

	rounds and assess performance after each round.	training phase has been completed.	
Evaluate Performance Metrics	Team discussed about the important measures for evaluating models, such as accuracy, precision, recall, and F1 score for classifying models.	Selected metrics for assessing model performance.	All team members should review and confirm the evaluation process. By the next meeting (30/03/2024).

Actions in brief:

- Trong Dat Hoang: Commence the testing of chosen models no later than Wednesday (27/09/2024).
- Henry Nguyen: Kick off the training of selected models by Wednesday (27/09/2024).
- Xuan Tuan Minh Nguyen: Get ready with tuning scripts and methodologies by Friday (29/09/2024).
- All team members: Assess and confirm the accuracy of evaluation metrics and processes before the next meeting.

Meeting closed at: 11:45AM – 23/09/2024

Next meeting time, date and location: 9:00AM – 30/09/2024. Meeting room 301, Library, Swinburne University of Technology, Hawthorn, VIC 3122.