THE 1st INTERNATIONAL ENGINEERING FESTIVAL

ARTIFICIAL INTELLIGENCE COMPETITION

PROJECT DOCUMENTATION

PROJECT TITLE:

ANALYSIS AND PREDICTION DASHBOARD FOR FLOODS IN MALAYSIA

AFFILIATE UNIVERSITY:

UNIVERSITI TEKNIKAL MALAYSIA MELAKA (UTeM)

PREPARED BY:

NO.	STUDENT'S NAME	MATRIC NO.	FACULTY
1	AQILAH FARHAH BINTI FAREZ EZAM	B032120043	FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
2	CLIFFSON ASHLEY ANAK DESMOND	B032110433	
3	NABILAH BINTI MOHD NOOR	B032120023	
4	NURSHAFIQAH BINTI KARIM	B032120037	

TABLE OF CONTENTS

1.0 INTRODUCTION

Flooding is an environmental hazard that occurs almost everywhere in the world. A flood is a body of water that overflows from the banks of a river, lake, or drainage system due to heavy rain, ice melting, high tide, and other factors. Flooding is one of the natural calamities that can cause significant property damage and even death. Most flood victims did not have enough time to save their belongings and crucial documents such as identification cards, birth certificates, and so forth. Historical records of floods have shown that flooding has an imminent impact on people's livelihoods, and it is unavoidable. Malaysia is not immune to flood calamities. In 1971, there was a flood tragedy that shocked all Malaysians involving 11 states in Malaysia. This tragedy has affected many cities to the point of paralysis. This is due to the persistent rain that lasted over a week. Because of this flood disaster, it can cause the destruction of residents' property. This is because a large flood can submerge houses, wash away goods, and damage items such as electrical items. Indirectly, it brings great loss to the population. Ironically, the effects of floods not only cause destruction and loss of property on a large scale but also result in the loss of human and animal lives.

The flood analysis and prediction dashboard for Malaysia is a tool for predicting future floods in Malaysia and analyzing flood tragedies in each state. It indirectly assists the community and authorities in preparing for the flood calamity. This analysis will make use of data from Malaysia's data.gov website from 2000 to 2010. This forecast is critical for the authorities to act quickly against the Malaysian community. This data analysis can deliver information in a fleeting time. This information is extremely useful to authorities in making sound decisions. By developing a real-time dashboard for the analysis of floods, it will show what is happening and what needs immediate attention and adjustment.

2.0 OBJECTIVES

- To generate a data analysis with a statistical dashboard for the use of natural disasters in floods in overseeing their data.
- To assist authorities in dealing with the problem of disasters in floods in the future.

3.0 PROBLEM STATEMENT

Data analysis is defined as a process of cleaning, transforming, and modeling data to discover useful information for business decision-making. However, without proper or wrongly used techniques in analyzing the obtained data, the outcomes might not be accurate and the chances of making poor decisions will increase. When dealing with natural disasters, especially floods, the decisions will involve lives, and loss of personal belongings and property damage by the disasters. Hence, with the advancement of artificial intelligence technologies, analysis and prediction dashboard for flood cases in Malaysia should be enhanced and improved. This will benefit the authorities such as the Ministry of Natural Resources, Environment and Climate Change, and Malaysia National Disaster Management Agency (NADMA) in dealing with natural disasters, especially flood circumstances. By having the enhanced and improved tools in analyzing and predicting natural disasters, standard precautions can also be improved and planned wisely, to reduce the effects of the disasters. Therefore, this will contribute to two Sustainable Development Goals of the United Nations (UN) international organization, which are Goal 11: Sustainable Cities and Communities, and Goal 13: Climate Action. In SDG 11, the aim is to make cities and human settlements inclusive, safe, resilient and sustainable (United Nation (UN), 2015). Meanwhile, for SDG 13, the aim is to take urgent action to combat climate change and its impacts (United Nation (UN), 2015). By retracing the advancement of the artificial intelligence technologies and the Sustainable Development Goals, it can be derived that Data Science can be used to improve humanity as it is the study of data to extract meaningful insights for business, especially in decision making for the management of natural disasters.

4.0 AIMS OF ARTIFICIAL INTELLIGENCE IN THE PROJECT

Artificial Intelligence (AI) is a field, which combines computer science and robust datasets to enable problem-solving. It contains human intelligence, such as perception, reasoning, learning, problem-solving, decision-making, and natural language processing. Moreover, this project uses Machine learning (ML) in implementing artificial intelligence techniques for Data Science domain to improve humanity.

The aim of this project is to propose data analysis for managing and maintaining the community's needs. The proposed system will be evaluated based on its capabilities, potential benefits, and challenges. The analysis will be based on the requirements and objectives that have been defined. The proposed intelligent system also includes components of data collection and processing, descriptive analytics, and predictive analytics.

4.1 Data Collection and Processing

The dataset that will be used is the dataset ‘rekod-banjir-2000-hingga-2010’. The dataset is from Malaysia's data.gov website covering the year 2000 to 2010. The site collects datasets from each state in Malaysia which is Perlis, Kedah, Pulau Pinang, Perak, Selangor, Kuala Lumpur, Negeri Sembilan, Melaka, Johor, Pahang, Terengganu, Kelantan, Sabah, and Sarawak.

SUMMARY OF FLOOD EVENT (2000 - 2010) BY RBMU FOR MALAYSIA														
RBMU No.	Item	2002	2012 Study (DNA)											Negeri
		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	TOTAL	
1	No. of Flood Event	1					1	1			2	1	5	Perlis
	River & Date of Flood						Sg. Perlis	Sg. Buluh			Sg. Jarum	Sg.		
							Sg. Arau	Sg. Alor			Sg.	Sg. Jarum		
								Sg. Arau			Sg. Buluh	Sg. Buluh		
											Sg.	Sg.		
												Sg.		
												Sg. Arau		
												Sg. Perlis		
												Sg.		
												Sg. Abi		
												Sg.		
												Sg.		
	RTB (Date of Completion)								10-Nov					
	Date of Worst Flood Event	No.v 2000												
	Area (km²) / ARI (Year)	18.36 / 10												
2	No. of Flood Event	1		2	1		1	1	5	1	4	1	16	Kedah
	River			Sg.	Sg.		Sg. Pdg.	Sg.	Sg.	Sg.	Sg.	Sg.		
				Sg. Bata /			Sg. Tg.	Sg. Pdg.	Sg. Pdg.	Sg.	Sg. Setar	Sg. Pdg.		
							Sg. Alor	Sg. Tg.	Sg. Pdg.	Sg. Pdg.	Sg. Pdg.	Sg. Bata		
							Sg.	Sg. Alor	Sg. Bata	Sg. Bkt.	Sg.	Sg. Anak		
							Sg. Baru	Sg.	Sg.	Sg. Anak	Sg. Anak	Sg.		
							Sg. Pdg.	Sg.	Sg. Badak	Sg. Nawa	Sg. Baru			
							Sg. Bata	Sg. Bata	Sg. Laka	Sg. Tg.	Sg. Pdg.			
							Sg.	Sg.		Sg. Bata	Sg. Bata			
										Sg. Alor	Sg.			
	RTB (Date of Completion)							26-May						
	Date of Worst Flood Event	Dec-00										30/10-3/11		
	Area (km²) / ARI (Year)	36.53 / 5										344.4 / -		
3	No. of Flood Event			0	2		0	0	1	3	2	0	8	
	River				Sg.				Sg.	Sg.	Sg.			
					Sg.				Sg.	Sg.	Sg.			

Figure 4.1 Dataset from data.gov.my

Figure 4.1 shows that the original dataset contains a lot of noise which needs to be cleaned and fixed before moving to the next phase. For example, the data combines two different data types on the same row such as river and date of the row that combine place name and date. Another noise is contained in the dataset which has too many null values which is hard to describe and analyze the data. The method to fix it is by using data cleaning. Data cleaning is a process of correcting errors or inconsistencies or restructuring data to make it easier to use.

For this dataset, the process cleaning has been deployed using the phyton language in Google Collab. Figure 4.2 below shows the dataset after the process of data cleaning has been successful, the dataset will be structured and doesn't contain a null value.

tempat	tarikhMula	tarikhTam	Tahun	Negeri
Sg. Kulim	17-Oct	18-Oct	2001	PulauPinang
Sg. Tok Sa	17-Oct	18-Oct	2001	PulauPinang
Sg. Maklu	17-Oct	18-Oct	2001	PulauPinang
Sg. Pertam	13-Jan	13-Jan	2001	PulauPinang
Sg. Kubang	13-Jan	13-Jan	2001	PulauPinang
Sg. Mengk	13-Jan	14-Jan	2001	PulauPinang
Sg. Ara Ku	13-Jan	13-Jan	2001	PulauPinang
Sg. Sintuk	13-Jan	13-Jan	2001	PulauPinang
Sg. Perai	13-Jan	13-Jan	2001	PulauPinang
Sg. Permat	13-Jan	14-Jan	2001	PulauPinang
Sg. Kilang	13-Jan	13-Jan	2001	PulauPinang
Sg. Macha	13-Jan	13-Jan	2001	PulauPinang
Sg. Macha	21-Jan	21-Jan	2001	PulauPinang
Sg. Junjon	13-Jan	15-Jan	2001	PulauPinang
Sg. Junjon	13-Jan	13-Jan	2001	PulauPinang
Sg. Junjon	21-Jan	21-Jan	2001	PulauPinang
Sg. Junjon	2-Oct	2-Oct	2001	PulauPinang
Sg. Bukit T	13-Jan	13-Jan	2001	PulauPinang
Sg. Pinang	27-Aug	27-Aug	2001	PulauPinang
Sg. Jeluto	2-Nov	2-Nov	2001	PulauPinang
Sg. Air Hit	2-Nov	2-Nov	2001	PulauPinang

Figure 4.2 Dataset that has been Cleaned

4.2 Descriptive Analytics

Descriptive analytics is one of the types of data analytics. It focuses on describing and summarizing historical data to provide insights and understanding of past events. It involves examining data to identify patterns, trends, and relationships between variables, and presenting the results in a clear and concise manner. Descriptive analytics can help organizations understand what has happened in the past and identify areas for improvement.

It allows the authorities to summarize and aggregate data to provide an overall understanding of the current situation and performance of the area of floods in Malaysia. Descriptive analytics can help the authorities to monitor the community's needs. The data analysis will be able to summarize and describe data by identifying patterns and trends within the data, which can be used to generate visualizations, dashboards, and identify key metrics to track performance.

4.3 Predictive Analytics

Predictive analytics is one of the types of data analytics. Predictive analytics can predict the possibilities of a particular event in the future. Predictive analytics is an important component in data science. It uses historical data to predict future trends and patterns, allowing the organization to proactively address potential issues or capitalize on opportunities. Predictive analytics can help the authorities to optimize their resources and improve their service to the community. Example, the predictive analytics on the total of floods history that happened, to predict the future of floods that can be happen.

5.0 DEPLOYMENT

Due to time constraints, the development of the dashboard for the analysis and prediction of floods in Malaysia to be a stand-alone system cannot be executed yet. Before being developed, the artificial intelligence model should be analyzed carefully and able to give the outmost accuracy during the data training and data testing phases when processing the obtained data. Technically, this is normal situation in Data Science domain, as a precaution step to prevent inaccurate results and unwanted outcomes.

6.0 TOOLS

Google Colaboratory is one of the tools being used to develop the artificial intelligence (AI) model, especially for analytical and predictive modeling. The open-source tool enables anyone with Internet access to experiment with machine learning and coding for artificial intelligence models. Hence, allowing a team to develop a model, anywhere, everywhere, and anytime. Additionally, it also can increase physical inactivity in the development of a model or project, allowing work productivity to have better improvement for the development team. The tool is also being integrated with Python libraries as well. As Python is the main programming language used by the development team to develop the AI model, using Google Colaboratory eases the analysis process of the model, together with the data cleaning, data manipulations, data training, and also data testing. On the other hand, in this model development, Looker tool is being used also. It is a cloud-based Business Intelligence (BI) tool that aids developers in exploring, sharing, and visualizing data that drive better business decisions. Besides, it allows anyone in the development team to analyze and find insights for the dataset quickly. This allowed the development team to have a better insight of the data in creating the analytic and predictive model.

7.0 PROGRAMMING LANGUAGE

The programming language being used in this project is Python, an interpreted, object-oriented, and high-level programming language with dynamic semantics. Nowadays, Python is being used to do data analysis, as it has a flexible integrated development environment (IDE), easy to learn, open-source, and well-supported with plenty of useful analytics libraries available. Specifically, in the Data Science area, when developing an analytic model or a predictive model, Python is the most convenient when dealing with big data, as it offers simplicity and readability for the users, especially programmers.

8.0 OUTPUTS

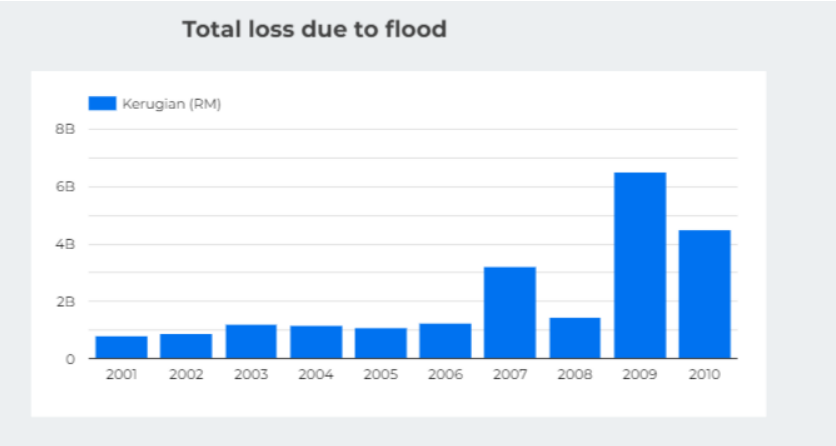

Figure 8.1 Total Loss Due to flood

Figure 8.1 shows the bar graph of total loss due to floods that happened in the year 2001 until 2010. From this, it shows that 2009 has the highest total loss due to flood. Through this total loss, the authorities can act by providing assistance to the community.

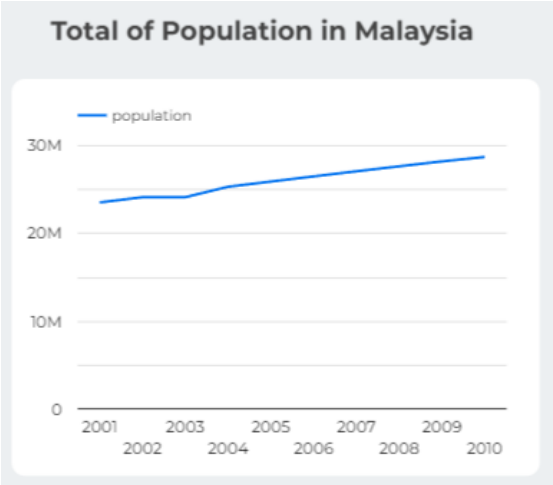

Figure 8.2 Total Population

Figure 8.2 shows the line graph of total loss of population in the year 2001 until 2010. From this, it shows that the population in the human population is increasing over time. It affects the number of flood victims involved in the disaster as well as the increase of the recovery cost.

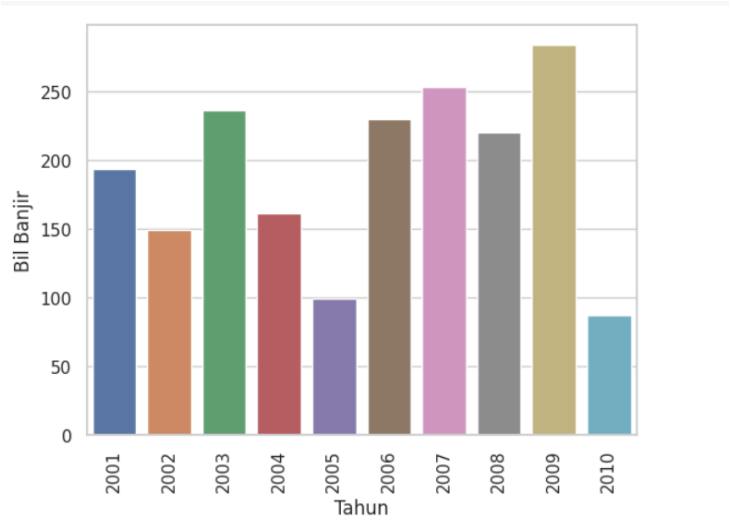

Figure 8.3 Flood frequency by year in Malaysia

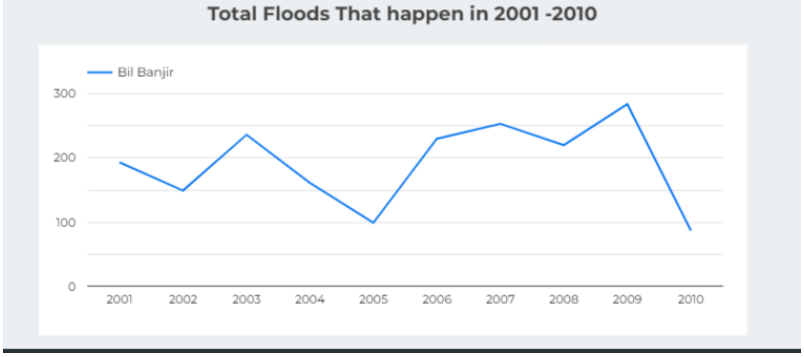

Figure 8.4 Total Floods from 2001 until 2010

Figure 8.3 and 8.4 shows the graph of flood frequency by year in Malaysia. From this, in 2009, floods occurred more often.

Figure 8. 5Analysis And Prediction Dashboard For Floods In Malaysia

Figure 8.5 shows Analysis And Prediction Dashboard For Floods In Malaysia using the Looker Data Studio. After done processing and run the coding in the google colab using the language phyton, we visualize the output in Looker to make it more clearly.

9.0 USAGE

Figure 9. 1 Analysis And Prediction Dashboard For Floods In Malaysia

Benefit from deploy this Analysis And Prediction Dashboard For Floods In Malaysia is it can help the authorities or organization to make preparation early before the floods happen such as the government can organize strategies in dealing with natural disasters that are floods in a more efficient and orderly manner. Initial preparations in terms of the number of rescuers on duty in that situation, budget preparations in buying equipment and food for flood victims.

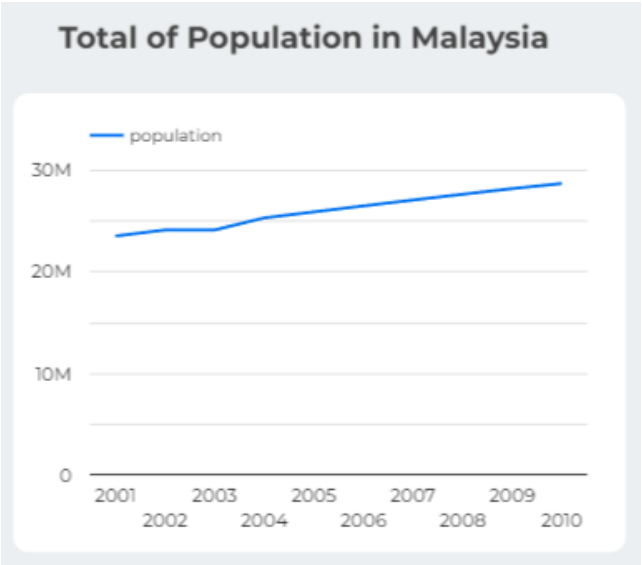

Figure 9.2 Total Floods from 2001 until 2010

Moreover, through the population data that is visualized in the graph world can help the government to plan well, that is, when the population is increasing, it will cause an area to have too dense population and indirectly when a disaster occurs, it will cause the need to rescue is getting higher. For example, the provision of a large number of lifeboats, accommodation places for flood victims also need to be prepared according to the number of flood victims and the need for wet and dry goods for flood victims can also be arranged more efficiently so that all flood victims and the situation when facing a disaster can be well managed.

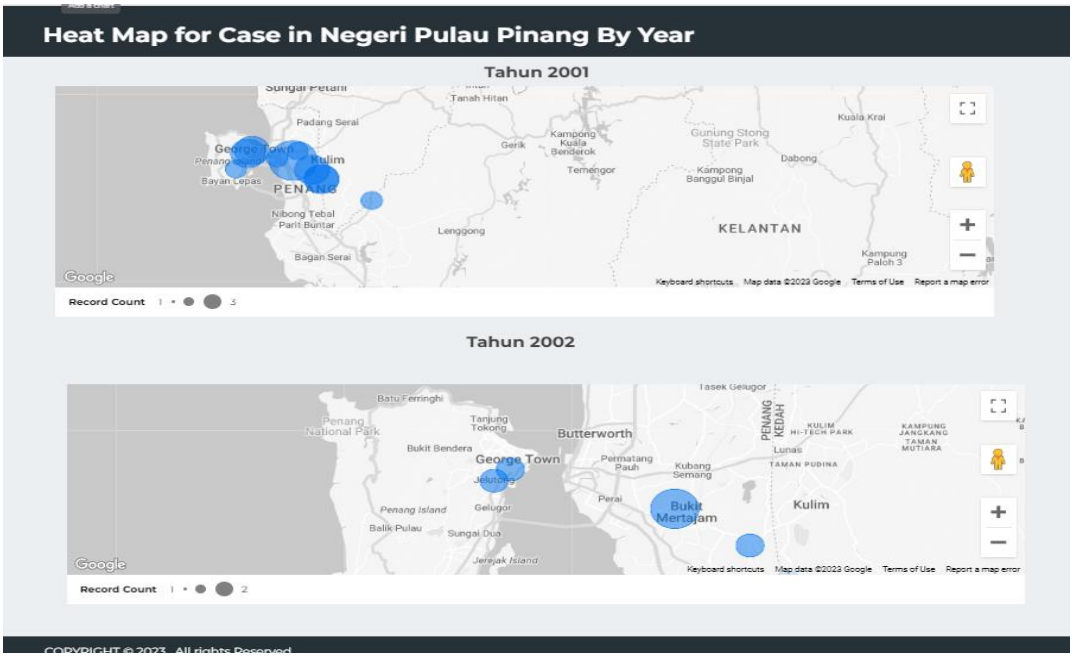

Figure 9.3 Heat Map For Case In Negeri Pulau Pinang By Year

In addition, Heat Map For Case which is data obtained from phyton and visualized in google locker studio dashboard can help the government to better see and know where the location of floods that often occur in each state.

Lastly, from this Analysis And Prediction Dashboard For Floods In Malaysia, it it is compatible with The Sustainable Development Goals (SDGs) asean such as Goaoal 13: Climate Action. In SDG 11, the aim is to make cities and human settlements inclusive, safe, resilient and sustainable (United Nation (UN), 2015) Goal 11: Sustainable Cities and Communities, and Goal 13: Climate Action. In SDG 11, the aim is to make cities and human settlements inclusive, safe, resilient and sustainable (United Nation (UN), 2015). Meanwhile, for SDG 13, the aim is to take urgent action to combat climate change and its impacts (United Nation (UN), 2015)

10.0 CONCLUSION

The ideas of developing something to analyze and obtaining insight from data is the main contribution for the growth of Data Science field, where artificial intelligence techniques were being applied optimally, especially in improving the life quality of the humanity. Hence, the things beyond human capability can be countered through the advancement of technologies. Analysis and prediction dashboard for floods in Malaysia is very important to ensure the well-being of the Malaysian community. With data visualization tools such as charts and graphs, the dashboard aims to provide a comprehensive overview of floods in Malaysia. This system provides a clear, up-to-date representation of what is and isn't working for users using a dashboard. They will spend less time since the reports are simple to grasp. The fact that users always have access to the information they require is very valuable. When everything is in one place, it is so much easier to analyze the data, spot trends, patterns, and take action on what we learn. The organization can perform better when they are fully aware of what is working and what needs to be improved.

11.0REFERENCES

Barkved, K. (2022, February 11). *The Difference Between Training Data vs. Test Data in Machine Learning*. Retrieved from obviously.ai logoobviously.ai logo: https://www.obviously.ai/post/the-difference-between-training-data-vs-test-data-in-machine-learning#:~:text=In%20machine%20learning%2C%20datasets%20are,known%20as%20the%20testing%20data

Ministy of Natural Resources, Environment and Climate Change . (2023). *River Water Level Data*. Retrieved from Water Level: https://publicinfobanjir.water.gov.my/aras-air/data-paras-air/?state=PNG&lang=en

United Nation (UN). (2015). *SUSTAINABLE DEVELOPMENT GOAL 11: SUSTAINABLE CITIES AND COMMUNITIES*. Retrieved from The Global Goals: https://www.globalgoals.org/goals/11-sustainable-cities-and-communities/

United Nation (UN). (2015). *SUSTAINABLE DEVELOPMENT GOAL 13: CLIMATE CHANGE*. Retrieved from The Global Goals: https://www.globalgoals.org/goals/13-climate-action/