Deliverable 1: Areas of Interest

Name: Tian Lan

Idea I – Malware Analysis and Detection Using Machine Learning Algorithms

- Question 1: What are the crucial features or patterns that can be extracted from malware samples to significantly improve the accuracy of malware detection?
- Question 2: Can a machine learning model be developed to dynamically adapt and identify newly emerged software as potential malware or benign applications?
- Question 3: How effectively can machine learning algorithms generalize and detect different variants of malware, ensuring comprehensive protection against evolving threats?

Business Use Case/ Potential Value:

By uncovering the key features and patterns that enhance the accuracy of malware detection, businesses and security professionals can develop robust defence mechanisms against malicious software. The potential target audience for this research includes cybersecurity professionals, software developers, system administrators, and organizations seeking to fortify their defence against malware. The outcomes of this study can empower these stakeholders with effective techniques, models, and strategies to bolster their security posture, minimize vulnerabilities, and protect their digital assets from potential breaches and disruptions.

Idea II - Predicting TikTok Trends Using Machine Learning Algorithms

- Question 1: What are the primary factors that significantly influence the emergence of TikTok video trends?
- Question 2: To what extent can machine learning algorithms accurately predict TikTok trends?
- Question 3: What is the potential viewership associated with each trending category on TikTok?

Business Use Case/ Potential Value:

By identifying the key factors contributing to the rise of TikTok video trends, businesses and content creators can enhance their understanding of the elements that resonate with the platform's user base. This knowledge can enable targeted content creation, strategic marketing campaigns, and improved audience engagement. The potential target audience for this research encompasses content creators, social media marketing teams, brand managers, and business strategists seeking to optimize their TikTok presence and leverage the platform's immense reach and influence. The outcomes of this study can empower these stakeholders to make data-driven decisions, enhance their brand visibility, and ultimately maximize their impact in the dynamic TikTok ecosystem.

Idea III - Fine-tuning a GPT Model for Improved Text Completion in a Specific Domain

- Question 1: How can we customize a pre-trained GPT model using specific domain data to improve text completion?
- Question 2: What effects do different fine-tuning approaches, such as data size, training time, and hyperparameter settings, have on the performance of the text completion model?
- Question 3: How does the fine-tuned GPT model compare to other text completion methods in terms of completion quality and relevance within the specific domain?

Business Use Case/ Potential Value:

The fine-tuning of a GPT model holds immense potential for revolutionizing language generation and streamlining processes. By customizing the model to specific domains, businesses can generate highly accurate and contextually relevant text completions, saving time and effort in content creation and customer support. This automation of tasks such as article writing, social media posts, and email responses enhances efficiency and enables the production of abundant, top-quality content. Additionally, the fine-tuned GPT model facilitates personalized recommendation systems, particularly valuable for CRM purposes, empowering e-commerce, streaming, and content platforms to offer tailored suggestions based on user preferences. Technology companies specializing in natural language processing, along with e-commerce platforms, customer support providers, and marketing agencies, are the potential target audience for this research, as they can leverage the advantages of the fine-tuned GPT model to enhance products, improve customer experiences, and streamline content creation and recommendation processes.

Idea IV - Food Shortage Prediction Using Data Science and Machine Learning

- Question 1: What are the primary factors influencing the supply chain from farm to table?
- Question 2: Is it possible to create a machine learning model that leverages past agricultural and climate data, along with other relevant factors, to predict future food shortages in specific regions?
- Question 3: What strategies can be employed to identify early indications of potential food shortages?

Business Use Case/ Potential Value:

The issue of food supply is a widespread and longstanding problem that has persisted for many years. However, by utilizing data science and machine learning techniques to anticipate and address food shortages, stakeholders can make well-informed choices, implement proactive measures, and optimize the allocation of resources. This allows them to ensure that individuals, particularly vulnerable populations, have reliable access to enough food that is safe, nutritious, and able to meet their dietary requirements. Ultimately, this approach aims to enable people to maintain healthy and active lives, while striving for universal food security.

Idea V – Personalized Weight Control using Data Science and Machine Learning

- Question 1: Can machine learning algorithms analyze personal health data, including factors like diet, exercise, sleep patterns, and medical history, to predict optimal weight control strategies for individuals?
- Question 2: How can wearable devices and sensor data be integrated with machine learning algorithms to track and provide real-time feedback on physical activity, calorie intake, and weight fluctuations?
- Question 3: Can machine learning models predict potential bottlenecks and challenges that individuals may face during weight control journeys, and provide targeted interventions to overcome them?

Business Use Case/ Potential Value:

Using data science and machine learning in weight management involves leveraging advanced analytics and algorithms to analyze individual health data and provide personalized insights and interventions. This approach enables precise prediction of weight trends, tailoring recommendations for diet, exercise, and behaviour modification, and real-time tracking of progress. The benefits include improved effectiveness of weight management strategies, personalized and data-driven interventions, enhanced engagement and adherence, and the ability to optimize outcomes by considering individual characteristics and preferences. Potential beneficiaries encompass weight loss applications, online platforms, healthcare providers, fitness centers, insurance companies, and other relevant stakeholders.