หน้าที่ 1

Hello everyone,  
Today, I’ll share the progress of my project titled **"Machine Learning Models for Head Injury Survival Prediction in Motorcycle Accidents."**  
I will cover data preparation, model selection, and initial evaluation results.

เฮลโล อีฟวรีวัน  
ทูเดย์ ไอล์ แชร์ เดอะ โพรเกรส ออฟ มาย โปรเจกท์ ไทเทิ่ลด์ **"แมชชีน เลิร์นนิ่ง โมเดล ฟอร์ เฮด อินเจอรี เซอร์ไววัล พรีดิกชั่น อิน โมเตอร์ไซเคิล แอคซิเดนต์ส"**  
ไอ วิล คัฟเวอร์ เดต้า เพรพเพอเรชั่น, โมเดล ซีเลคชั่น, แอนด์ อิเนียเชียล อีวาลูเอชั่น รีซัลต์ส

หน้าที่ 2

According to the plan for Week 7, we are still in the phase of refining the models and have started writing the manuscript.

แอ็คคอร์ดิง ทู เดอะ แพลน ฟอร์ วีค เซเว่น, วี อาร์ สติลล์ อิน เดอะ เฟส ออฟ รีไฟน์นิง เดอะ โมเดลส์ แอนด์ แฮฟ สตาร์ทเท็ด ไรท์ติ้ง เดอะ แมนูสคริปต์

หน้าที่ 3

In the next step, we will modify the data in the **"Dead"** column. Originally, the value **1 represented deceased** and **0 represented survived**.  
However, since our model focuses on predicting the probability of **survival**, we have reversed these values as follows:

* **1 now indicates survival**
* **0 now indicates death**  
  This change will help make the model training and interpretation clearer and better aligned with the research objectives.

อิน เดอะ เน็กซ์ สเต็ป, วี วิล โมดิไฟ เดอะ เดต้า อิน เดอะ **"เดด"** คอลัมน์  
ออริจินัลลี่, เดอะ แวลู วัน รีเพรเซนต์เต็ด ดิเซสด์ แอนด์ ซีโร รีเพรเซนต์เต็ด เซอร์ไวฟด์  
ฮาวเอเวอร์, ซินซ์ เอา โมเดล โฟกัส ออน พรีดิกท์ เดอะ โพรบาบิลิตี้ ออฟ เซอร์ไววัล, วี แฮฟ รีเวิร์สด เดส แวลูส แอส ฟอลโลว์ส:

* วัน นาว อินดิเคตส์ เซอร์ไวฟัล
* ซีโร นาว อินดิเคตส์ เดธ  
  ดิส เชนจ์ วิล เฮลพ์ เมค โมเดล เทรนนิง แอนด์ อินเตอร์พริตเทชัน เคลียร์ แอนด์ แมตช์ วิธ เดอะ รีเสิร์ช ออบเจกทีฟส์

หน้าที่ 4

Next, we will show the data **before and after the modification** of the values in the **"Dead"** column.  
On the left side, the original data is displayed, where **1 indicates deceased** and **0 indicates survived**.  
On the right side, the modified data is shown, where the values have been reversed:

* **1 now represents survived**
* **0 now represents deceased**  
  This change aligns with the model’s focus on predicting survival.

เน็กซ์, วี วิล โชว์ เดอะ เดต้า **บีฟอร์ แอนด์ แอฟเตอร์** เดอะ โมดิฟิเคชัน ออฟ เดอะ แวลูส อิน เดอะ **"เดด"** คอลัมน์  
ออน เดอะ เลฟท์ ไซด์, เดอะ ออริจินัล เดต้า อิส ดิสเพลย์ด, แวร์ **วัน อินดิเคตส์ ดิเซสด์** แอนด์ **ซีโร อินดิเคตส์ เซอร์ไวฟด์**  
ออน เดอะ ไรท์ ไซด์, เดอะ โมดิฟายด์ เดต้า อิส โชว์น, แวร์ เดอะ แวลูส แฮฟ บีน รีเวิร์สด:

* **วัน นาว รีเพรเซนต์ เซอร์ไวฟด์**
* **ซีโร นาว รีเพรเซนต์ ดิเซสด์**  
  ดิส เชนจ์ อไลน์ส วิธ เดอะ โมเดลส โฟกัส ออน พรีดิกท์ิง เซอร์ไววัล

หน้าที่ 5

This page presents the two most accurate models selected from last week’s experiments for comparison during the refinement phase.

The images shown here display results based on the original data, prior to any modifications.

ดิส เพจ เพรสเซินท์ส เดอะ ทู โมส แอคเคอร์เรท โมเดลส ซีเล็คเต็ด ฟรอม ลาสต์ วีคส์ เอกซ์เพอริเมนต์ส ฟอร์ คอมแพริสัน ดัวริง เดอะ รีไฟน์เมนท์ เฟส

เดอะ อิมเมจเซส โชว์น เฮียร์ ดิสเพลย์ รีซัลต์ส เบสด์ ออน เดอะ ออริจินัล เดต้า, ไพรเออร์ ทู เอนี โมดิฟิเคชันส

หน้าที่ 6

The next slide presents data improvement using **Method 1**, focusing on addressing the problem of **imbalanced data**.

This step utilizes **SMOTE (Synthetic Minority Over-sampling Technique)**, which generates synthetic samples for the minority class by sampling nearby data points, thereby increasing the number of samples in the **deceased** group, which is smaller than the survived group.

After balancing the data, this dataset is used with the **two selected models** for comparing their performance under balanced data conditions.

เดอะ เน็กซ์ สไลด์ เพรสเซินท์ส เดต้า อิมพรูฟเมนต์ ยูซิง **เมธ็อด วัน**, โฟกัสซิ่ง ออน แอดเดรสซิง เดอะ ปรอบเล็ม ออฟ **อิมบาลานซ์ด เดต้า**

ดิส สเต็ป ยูทิลไลซ์ส **SMOTE (ซินเธทิค ไมโนริตี้ โอเวอร์-แซมพลิง เทคนิค)** วิช เจเนอเรทส ซินเธทิค แซมเพิลส ฟอร์ เดอะ ไมโนริตี้ คลาส บาย แซมพลิง เนียร์บาย เดต้า พอยต์ส, แดร์บาย อินครีสซิง เดอะ นัมเบอร์ ออฟ แซมเพิลส อิน เดอะ **ดิเซสด์** กรุ๊ป, วิช อิส สมอลเลอร์ แธน เดอะ เซอร์ไวฟด์ กรุ๊ป

แอฟเตอร์ บาลานซิง เดอะ เดต้า, ดิส เดต้าเสท อิส ยูสด์ วิธ เดอะ **ทู ซีเล็คเต็ด โมเดลส** ฟอร์ คอมแพริ่ง แธร์ เพอร์ฟอร์แมนซ์ อันเดอร์ บาลานซ์ด เดต้า คอนดิชันส์

หน้าที่ 7

Method 2 uses **Stratified K-Fold Cross-Validation**, a technique that splits the data into **K folds** while maintaining the **class proportions** close to the original dataset.

The model is trained **K times**, each time using a different fold as the test set, and the results are averaged.

This method provides a more **accurate and fair** evaluation, especially when dealing with imbalanced data.

เมธ็อด ทู ยูซส **สแทรทิไฟด์ เค-โฟลด์ ครอส-แวลิเดชัน**, อะ เทคนิค แธท สปลิทส เดอะ เดต้า อิ้นทู **เค โฟลด์ส** ไวล ไมเทนนนิ่ง เดอะ **คลาส โพรพอร์ชันส** โคลส ทู เดอะ ออริจินัล เดต้าเสท

เดอะ โมเดล อิส เทรนด **เค ไทม์ส**, อีช ไทม์ ยูซิง อะ ดิฟเฟอเรนท์ โฟลด์ แอส เดอะ เทสต์ เซท, แอนด์ เดอะ รีซัลต์ส อาร์ แอฟเวอริจด์

ดิส เมธ็อด โพรไวด์ส อะ มอร์ **แอคเคอร์เรท แอนด์ แฟร์** อีวาลูเอชัน, เอสเปเชียลลี เว็น ดิลลิ่ง วิธ อิมบาลานซ์ด เดต้า

หน้าที่ 8

This slide shows the results from using **Stratified K-Fold Cross-Validation** to compare the **Random Forest** and **Logistic Regression** models.

Both models achieved over **94% Accuracy** and **F1-score**, indicating strong overall performance.

* Logistic Regression has the highest **Recall** for the survived class (class 1) at 97.2%.
* Random Forest has a higher **Recall** for the deceased class (class 0) at 44.0% compared to 39.5%.
* The **AUC** of Random Forest (0.9019) is slightly higher than Logistic Regression (0.8985).

In summary, Logistic Regression excels at predicting survival, while Random Forest better distinguishes between classes overall.

ดิส สไลด์ โชว์ เดอะ รีซัลต์ส ฟรอม ยูซิง **สแทรทิไฟด์ เค-โฟลด์ ครอส-แวลิเดชัน** ทู คอมแพร์ เดอะ **แรนดอม ฟอเรสต์** แอนด์ **ลอจิสติก รีเกรสชัน** โมเดลส

โบธ โมเดลส แอชีฟด์ โอเวอร์ **94% แอคคิวราซี** แอนด์ **เอฟวัน-สกอร์**, อินดิเคติง สตรอง โอเวอร์ออล เพอร์ฟอร์แมนซ์

* ลอจิสติก รีเกรสชัน แฮส เดอะ ไฮเอสท์ **รีคอล** ฟอร์ เดอะ เซอร์ไวฟด์ คลาส (คลาส วัน) แอท 97.2%
* แรนดอม ฟอเรสต์ แฮส อะ ไฮเออร์ **รีคอล** ฟอร์ เดอะ ดิเซสด์ คลาส (คลาส ซีโร) แอท 44.0% คอมแพร์ด ทู 39.5%
* เดอะ **เอยูซี** ออฟ แรนดอม ฟอเรสต์ (0.9019) อิส สไลต์ลี ไฮเออร์ แธน ลอจิสติก รีเกรสชัน (0.8985)

อิน ซัมเมอรี, ลอจิสติก รีเกรสชัน เอ็กเซลแอต พรีดิกท์ิง เซอร์ไววัล, ไวล แรนดอม ฟอเรสต์ เบ็ทเทอร์ ดิสทิงกวิช บีทวีน คลาสเซส โอเวอร์ออล

From the two methods tested, **Stratified K-Fold Cross-Validation** demonstrates better performance than **SMOTE (Synthetic Minority Over-sampling Technique).**

ฟรอม เดอะ ทู เมธ็อดส เทสเต็ด, **สแทรทิไฟด์ เค-โฟลด์ ครอส-แวลิเดชัน** เดมอนสเตรทส เบ็ทเทอร์ เพอร์ฟอร์แมนซ์ แธน **SMOTE (ซินเธทิค ไมโนริตี้ โอเวอร์-แซมพลิง เทคนิค)**

I have started researching how to write the manuscript and have begun drafting a small portion of it.

ไอ แฮฟ สตาร์ทเท็ด รีเสิร์ชชิง ฮาว ทู ไรท์ เดอะ แมนูสคริปต์ แอนด์ แฮฟ บีกัน ดราฟทิง อะ สมอลล์ พอร์ชัน ออฟ อิท