



# STROKE PREDICTION

COMP6577 - Machine Learning - LC01

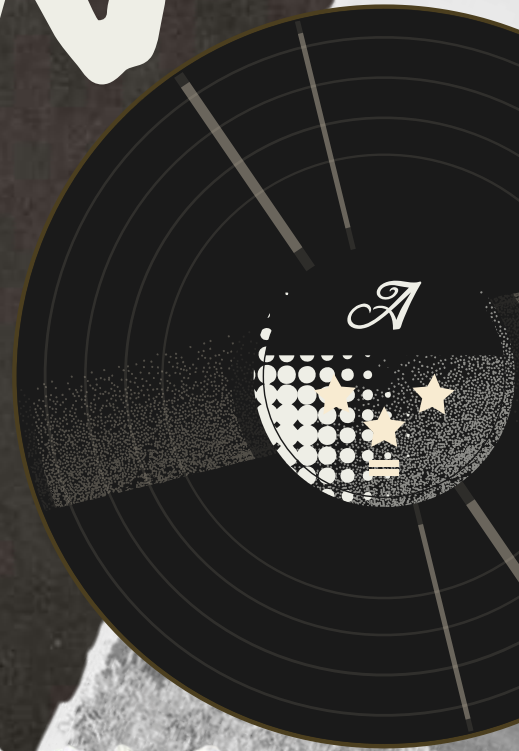


2301853136 - REINALDY SUKAMTO

2301855154 - AXEL LIE

2301937173 - DEAN ANANDA RAMADHAN

2301920941 - KELLY CORNELYA KESUMA

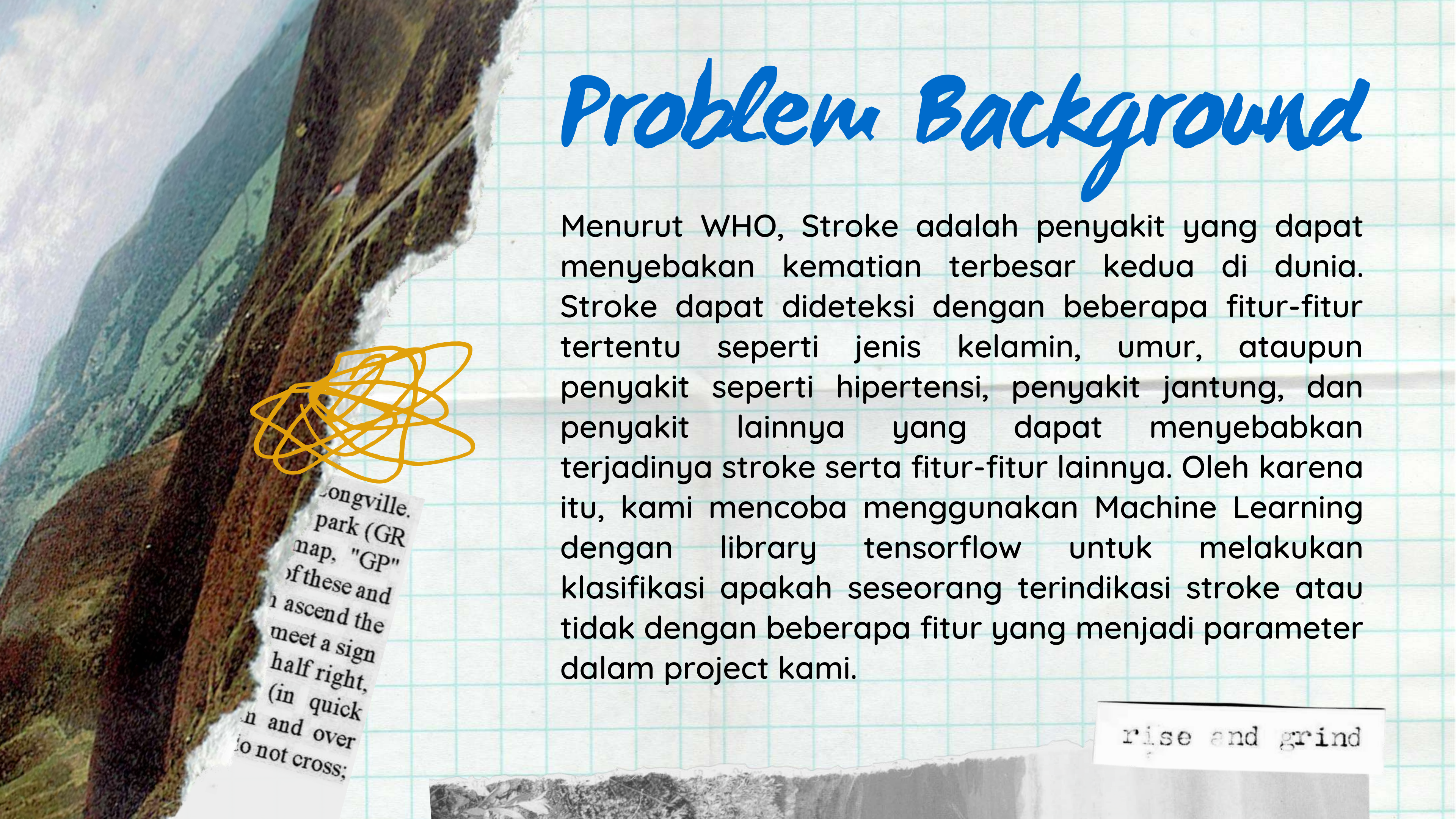


perfect



# Problem Background

Menurut WHO, Stroke adalah penyakit yang dapat menyebabkan kematian terbesar kedua di dunia. Stroke dapat dideteksi dengan beberapa fitur-fitur tertentu seperti jenis kelamin, umur, ataupun penyakit seperti hipertensi, penyakit jantung, dan penyakit lainnya yang dapat menyebabkan terjadinya stroke serta fitur-fitur lainnya. Oleh karena itu, kami mencoba menggunakan Machine Learning dengan library tensorflow untuk melakukan klasifikasi apakah seseorang terindikasi stroke atau tidak dengan beberapa fitur yang menjadi parameter dalam project kami.



Longville.  
park (GR  
nap, "GP"  
of these and  
ascend the  
meet a sign  
half right,  
(in quick  
n and over  
o not cross;

rise and grind



**PURPOSE :** Memprediksi apakah seseorang mengidap penyakit stroke atau tidak.

## BENEFITS

### PREVENTION

USER DAPAT MENCEGAH KEMATIAN  
YANG DIAKIBATKAN OLEH STROKE  
DENGAN MENGETAHUI LEBIH DINI  
APAKAH DIRINYA MENGIDAP STROKE  
ATAU TIDAK.

### SELF ANALYSIS

USER DAPAT MELAKUKAN  
PENGECEKAN SECARA MANDIRI  
(TANPA BANTUAN AHLI / DOKTER)  
APAKAH DIRINYA DINYATAKAN  
MENGIDAP STROKE ATAU TIDAK OLEH  
SYSTEM.

### QUICK TREATMENT

USER DAPAT MENERIMA HASIL  
PENGECEKAN SECARA CEPAT.



# LITERATURE REVIEW - 1

## HOW TO SOLVE IMBALANCED DATASET

Dari eksperimen yang dilakukan pada pertemuan sebelumnya, permasalahan yang dihadapi adalah **imbalanced dataset** dimana data non-stroke lebih banyak dari stroke. Dari jurnal (Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G, 2017), salah satu metode yang digunakan untuk mengatasi masalah **imbalanced dataset** adalah metode oversampling dengan menggunakan **SMOTE**. Hal ini menjadi **fokus utama** kelompok kami, dikarenakan dataset yang kami gunakan memiliki masalah berupa **imbalanced dataset**.

# LITERATURE REVIEW - 2

## SYNTETIC MINORITY OVERSAMPLING TECHNIQUE (SMOTE)

Berdasarkan paper (Chawla, Boyer, Hall, Kegelmeyer, 2002) SMOTE merupakan resampling jenis over-sampling dengan strategi cluster-based untuk mengatasi imbalanced data. SMOTE akan membuat data sintesis dari minority class berdasarkan feature space dari data tersebut. Dengan menggunakan K Nearest Neighbor, SMOTE dapat menentukan cluster yang didapatkan dan membuat data di cluster-cluster tersebut.

# LITERATURE REVIEW - 3

## BINARY CLASSIFICATION METHODS

Dari beberapa paper yang kita baca (Kotsiantis, 2007; Kumari, Srivastava, 2017), dikatakan bahwa terdapat beberapa method yang umum digunakan dalam melakukan **Binary Classification**. Dari beberapa pilihan method yang ada, kami memilih beberapa method yang telah kami pahami, diantaranya adalah **Support Vector Machine (SVM)**, **Feedforward Neural Network (FNN)**, **Decision Tree**, dan juga **Logistic Regression**.



# LITERATURE REVIEW

## REFERENCES

- Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. Expert Systems with Applications, 73, 220-239.
- Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, 160(1), 3-24.
- Kumari, R., & Srivastava, S. K. (2017). Machine learning: A review on binary classification. International Journal of Computer Applications, 160(7).
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.



# Methods

Kelompok kami telah mencoba menggunakan beberapa metode seperti:

1. Support Vector Machine Classifier
2. Neural Network - Feed Forward Neural Network
3. Decision Tree
4. Logistic Regression



## EXPERIMENT RESULT - PAGE 1

Untuk mengatasi permasalahan pada skewed dataset, pertama-tama kelompok kami menggunakan metode resampling. Namun, setelah dilakukan training dan evaluasi ditemukan bahwa model masih overfit terhadap dataset. Lalu, kelompok kami mengasumsikan bahwa hal ini disebabkan oleh metode resampling yang kurang efektif. Lalu, setelah melakukan literature review, kelompok kami mencoba untuk menggunakan SMOTE (Syntetic Minority Oversampling Technique). Dengan melakukan sedikit percobaan menggunakan SMOTE, ternyata permasalahan mengenai model yang overfitting dapat terselesaikan.



# RESULT BEFORE & AFTER USING SMOTE

## FEEDFORWARD NEURAL NETWORK

### BEFORE

- Accuracy: 88%
- Precision: 14%
- Recall: 20%
- F1: 16%

### AFTER

- Accuracy: 88%
- Precision: 87%
- Recall: 88%
- F1: 87%



## EXPERIMENT RESULT - PAGE 2

### *Support Vector Machine (SVM)*

- Accuracy: 88%
- Precision: 87%
- Recall: 88%
- F1: 87%

### *Feedforward Neural Network (FNN)*

- Accuracy: 88%
- Precision: 87%
- Recall: 88%
- F1: 87%



## EXPERIMENT RESULT - PAGE 3

### *Decision Tree*

- Accuracy: 94%
- Precision: 95%
- Recall: 94%
- F1: 94%

### *Logistic Regression*

- Accuracy: 77%
- Precision: 74%
- Recall: 80%
- F1: 77%



# Future Works

- PRODUCTION-GRADE ALGORITHMS (GRADIENT BOOST, RANDOM FOREST, OTHER ENSEMBLE LEARNING)
- DEEPER EXPLORATION
- SOLVE IMBALANCED FEATURE





THANK

YOU

