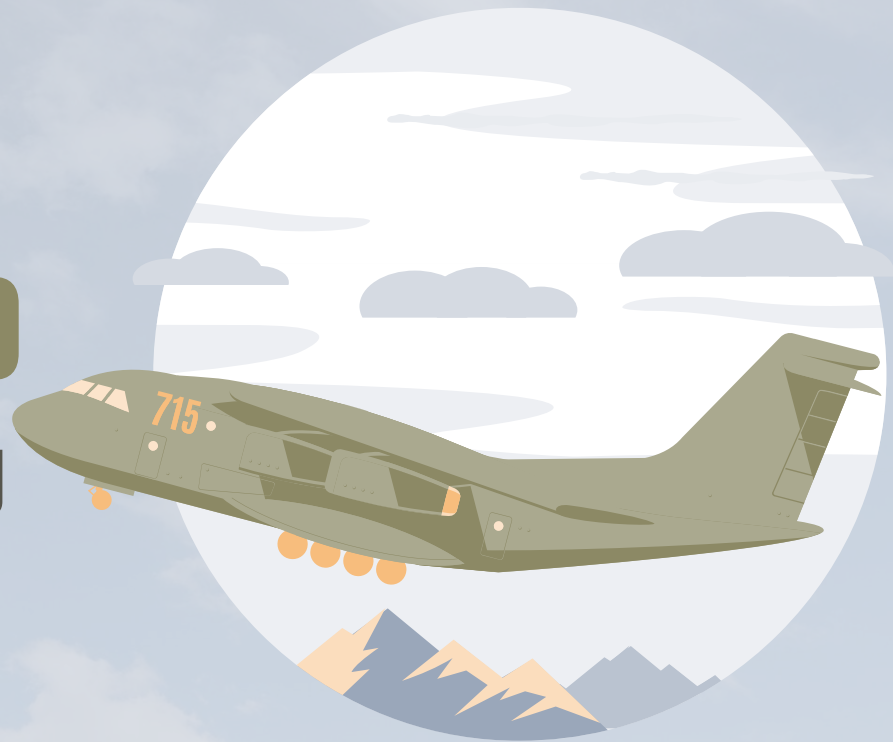


SPARK MLlib

Spark Machine Learning



Haarisah - Ony - Reynaldi

Table of contents

-01-

Pendahuluan

-02-

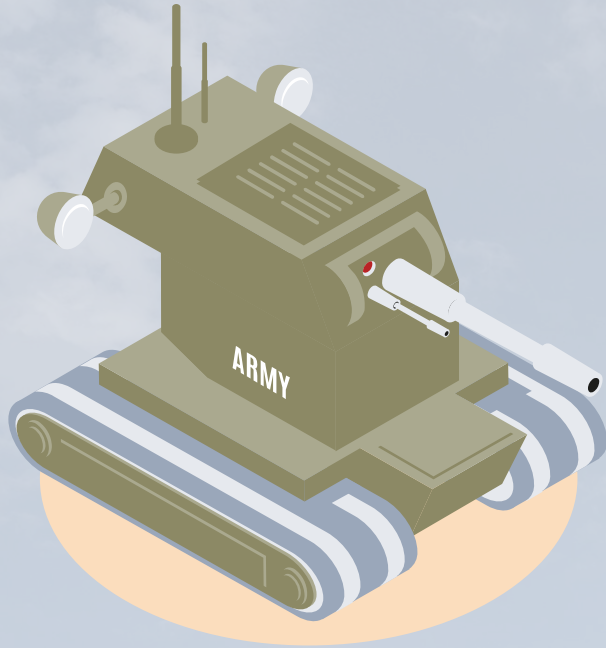
Flowchart

-03-

Rencana Pengolahan

-04-

Hasil Yang Diharapkan



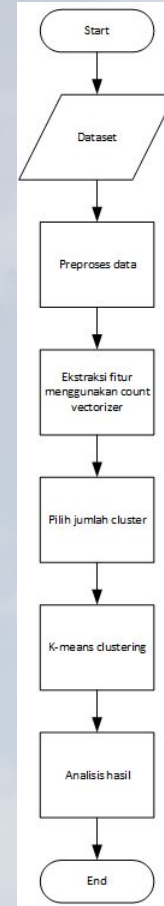
Pendahuluan

Dalam pengolahan data untuk menentukan faktor penyebab kecelakaan pesawat, kami berencana untuk menggunakan metode **k-means clustering** untuk membagi data ke dalam kluster tertentu yang memiliki kesamaan.

Flowchart

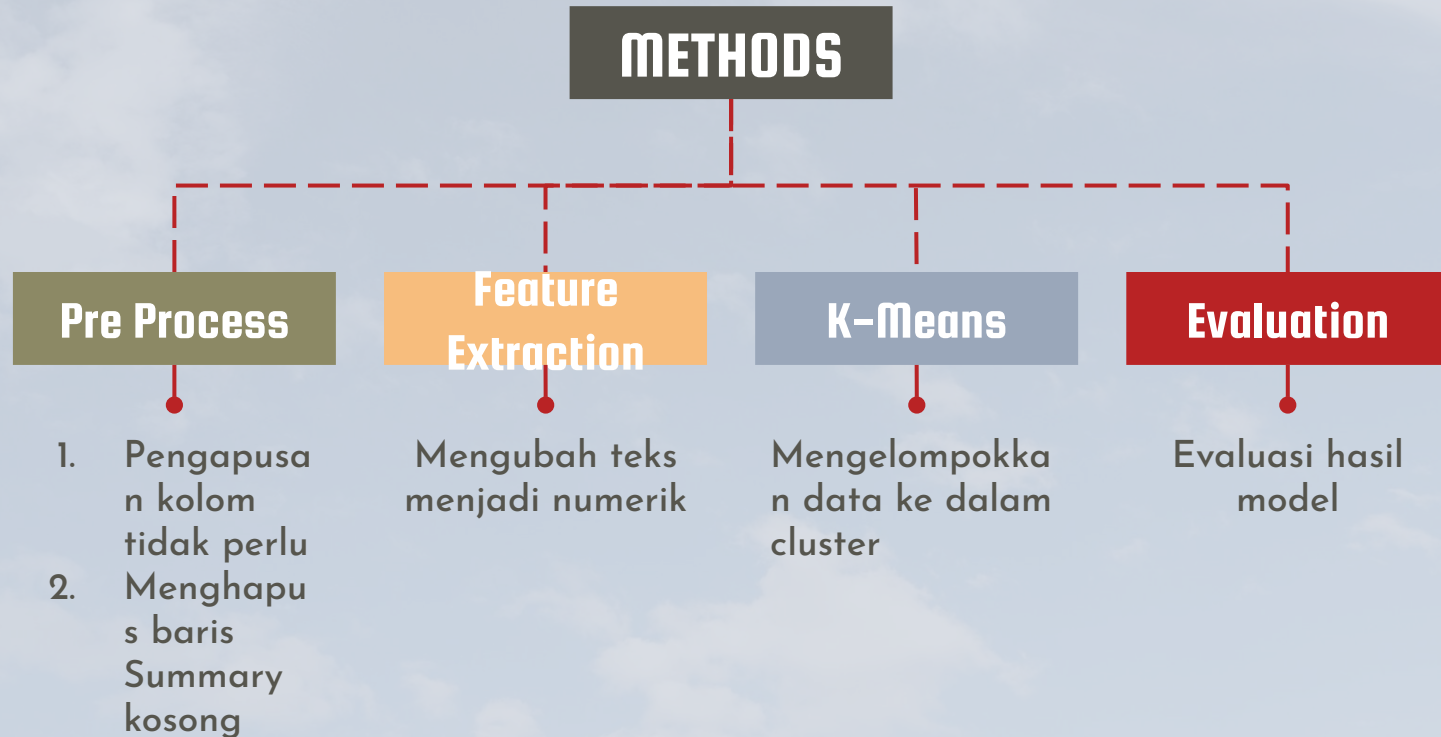
Bagan Alir yang terdiri dari 5 tahapan (setelah data didapat) :

1. Pre-proses Data
2. Ekstraksi Fitur Menggunakan Count Vectorizer
3. Pilih Jumlah Cluster
4. K-means Clustering
5. Analisis Hasil



Flowchart : [Link Flowchart](#)

Rencana Pengolahan



Hasil Yang Diharapkan



Jarak

Jarak antara titik data dengan centroid cluster harus berdekatan, sehingga menunjukkan bahwa suatu cluster memiliki kesamaan yang tinggi



Kualitas

Setiap cluster tidak tumpang tindih dan terpisah satu sama yang lain (memiliki karakter yang konsisten)



Hasil

Cluster memiliki arti yang jelas dan dapat diinterpretasikan



Terimakasih