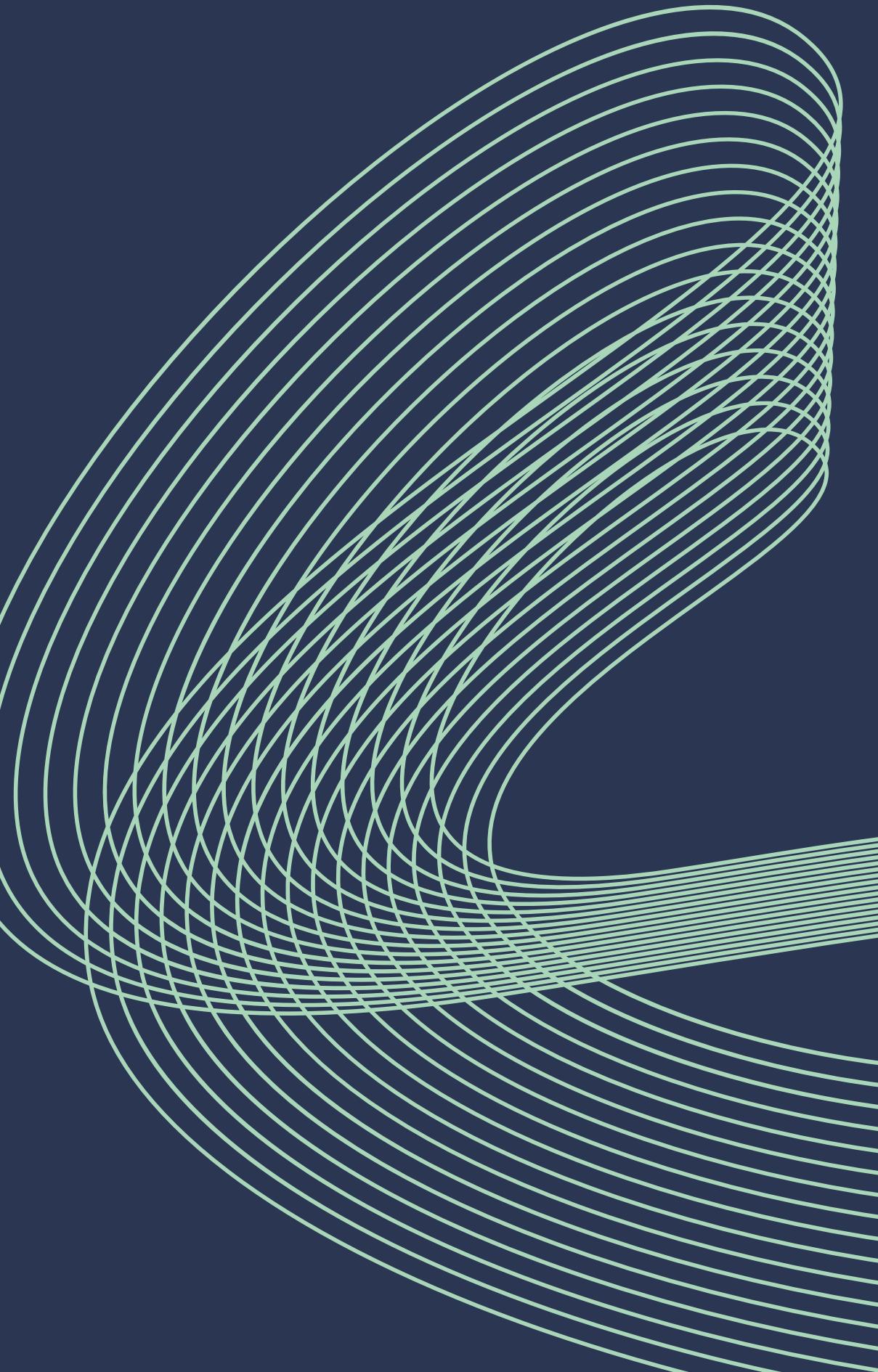


# Klasifikasi Saham IDX Kepada Listing Board Berdasarkan 5 Faktor Utama

Kelompok 5





# OUR TEAM

- Arya Jayavardhana - 00000060218
- Leonardo - 00000060177
- Steven Marcelino Tandiono - 00000059949
- Vinka Bella - 00000061043

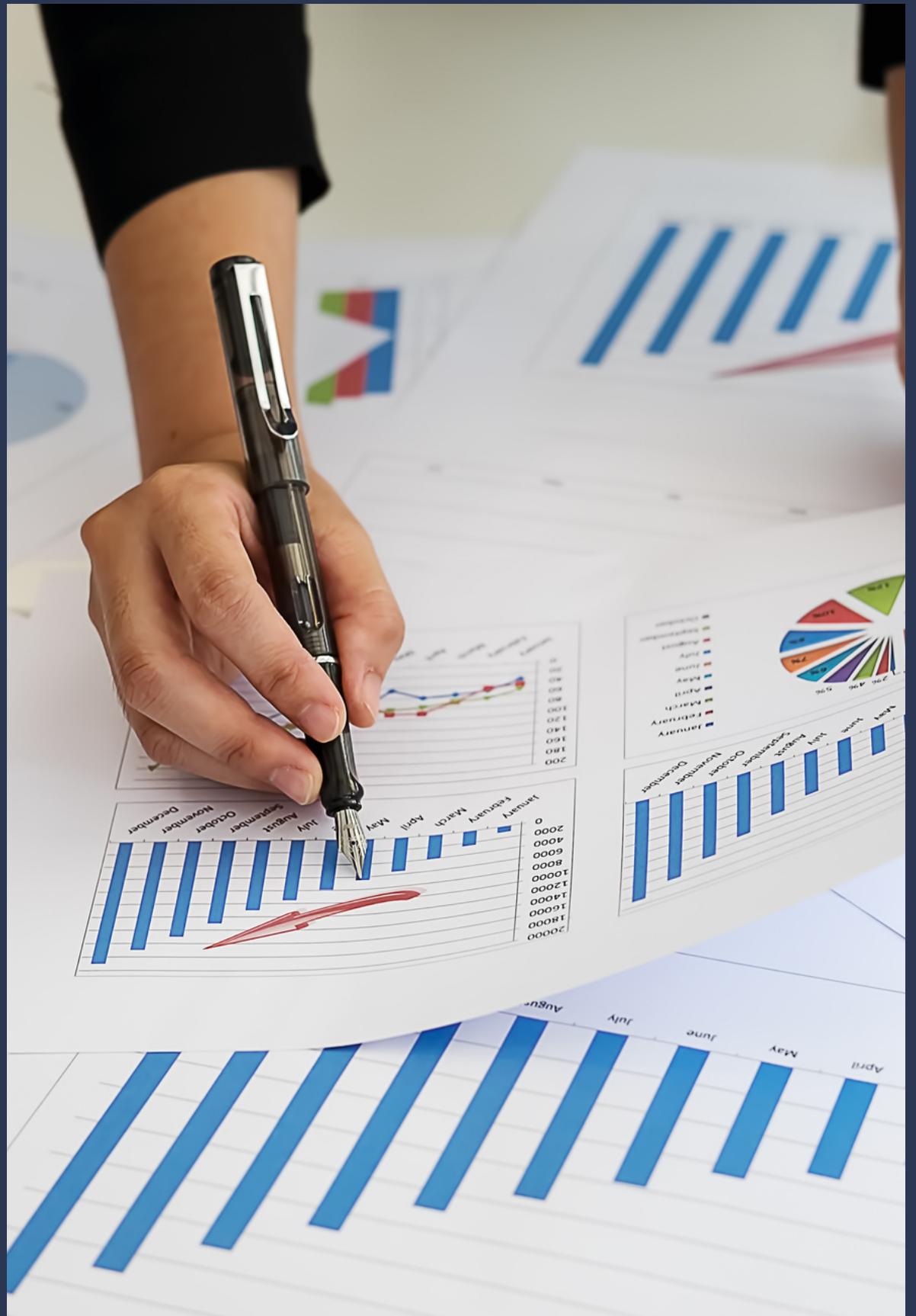
# Table of Content

Latar Belakang	1.	Telaah Literatur	6.
Rumusan Masalah	2.	Contoh Literatur	7.
Batasan Masalah	3	Metodologi	8.
Tujuan Penelitian	4.	Hasil dan Diskusi	9.
Manfaat Penelitian	5.	Kesimpulan	10.

# Latar Belakang

Listing board merupakan sebuah platform dimana saham sebuah perusahaan dapat dikategorikan dan diperdagangkan secara publik pada pasar paham. Dalam mengklasifikasikan suatu saham perusahaan kepada listing board yang terbagi menjadi tiga yaitu utama, pengembangan, dan akselerasi membutuhkan setidaknya beberapa faktor utama.

Faktor yang dibutuhkan masih terbilang cukup banyak dan akan memakan cukup banyak waktu bisa minggu sampai dengan bulan sehingga masih tidak terlalu efisien. Dalam hal ini, projek ini dilakukan untuk mengetahui apakah dengan hanya memakai lima faktor yaitu shares, market cap, value, frequency, operational lifetime masih dapat mengklasifikasikan saham ke dalam listing board dengan hasil yang cepat dan akurat.



# Rumusan Masalah

- Faktor manakah yang paling berpengaruh dalam melakukan klasifikasi saham kepada listing board
- Apakah 5 faktor mampu untuk mengklasifikasikan saham kepada listing board?
- Algoritma manakah yang mempunyai performa dan hasil yang lebih baik?



# Tujuan Penelitian

1. Menentukan atau mengidentifikasi faktor-faktor penentu yang mempengaruhi klasifikasi saham kepada listing board.
2. Menentukan apakah 5 faktor tersebut cukup untuk melakukan klasifikasi saham.
3. Menentukan algoritma yang dapat menghasilkan peforma lebih baik.



# Batasan Masalah

1. Hanya dapat melakukan klasifikasi terhadap saham yang terdaftar pada IDX saja.
2. Hanya dapat melakukan klasifikasi ke dalam listing board saham IDX.
3. Tidak dapat mempertimbangkan faktor eksternal yang dapat terjadi kepada bursa saham.



# Manfaat Penelitian

## Manfaat Teoritis

1. Mengidentifikasi faktor-faktor penting dalam klasifikasi saham.
2. Meningkatkan pemahaman tentang faktor penentu saham kedalam listing board.

## Manfaat Praktis

1. Meningkatkan efisiensi bursa saham dalam melakukan klasifikasi saham kedalam 3 klasifikasi listing board.
2. Membantu investor dalam melakukan pengambilan keputusan investasi saham.



# Telaah Literature

## K-Nearest Neighbor

K-Nearest Neighbor (KNN) adalah suatu metode yang dapat digunakan untuk mengklasifikasikan data uji baru menurut data pembelajarannya, yang berdasarkan mayoritas nilai K tetangga dengan jarak terdekatnya[2]. Singkatnya, tujuan dari pemodelan KNN yaitu untuk mengklasifikasikan data uji baru berdasarkan mayoritas K tetangga terdekat dari data pelatihan. Dalam menentukan nilai K terbaik pada pemodelan tergantung pada sifat data yang digunakan.

Dengan nilai K yang tinggi akan mengurangi efek noise atau outlier dalam data, akan tetapi juga menciptakan risiko generalisasi atau underfitting berlebih yang membuat proses klasifikasi menjadi kabur. Sedangkan nilai K yang rendah akan menangkap detail yang lebih baik, namun juga menciptakan efek noise atau outlier yang tinggi pada data.

# Contoh Literature

Pada algoritma KNN untuk mencari dekat atau jauhnya tetangga terdekat, pada umumnya dapat dihitung dengan menggunakan jarak Euclidean. Adapun rumus tersebut dapat dilihat seperti pada persamaan 1, yaitu [5]:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Terdapat serangkaian tahapan untuk menghitung algoritma atau metode dari K-Nearest Neighbors, yakni[6]:

1. Menentukan nilai parameter K atau jumlah tetangga terdekat.
2. Menghitung jarak antara data latih dan data uji menggunakan perhitungan rumus jarak euclidean.
3. Mengurutkan hasil perhitungan secara ascending.
4. Mengklasifikasikan data uji baru berdasarkan label terbanyak dari nilai K.

# Telaah Literature

## Decision Tree

Decision Tree atau pohon keputusan adalah metode yang digunakan untuk pengklasifikasian maupun regresi. Metode Decision tree merupakan alat pembuat keputusan dengan menggunakan model keputusan yang memiliki struktur seperti pohon. Layaknya sebuah pohon, Decision tree memiliki tiga bagian utama yaitu simpul akar, cabang, dan simpul daun. Dimana simpul akar mewakili fitur atau atribut, setiap cabang mewakili aturan keputusan, dan sedangkan simpul daun mewakili label tiap kelas [3]. Tujuan dari metode ini adalah memprediksi target berdasarkan aturan keputusan yang dibentuk dari dataset dan fitur terkait [4].



# Contoh Literature

Dalam algoritma Decision Tree, perlu dilakukannya penentuan atribut yang terbaik untuk simpul akar dan simpul daunnya. Dimana permasalahan ini dapat diselesaikan dengan menggunakan teknik perhitungan seleksi atribut, seperti entropi dan information gain. Entropi merupakan menghitung keacakan atau ketidakmurnian dalam data, semakin tinggi entropi berarti data semakin acak. Rumus dari perhitungan entropi terlihat seperti pada persamaan, yaitu:

$$Entropy(S) = \sum_{i=1}^c -P_i \log_2 P_i$$

Sedangkan information gain merupakan ukuran perubahan entropi untuk menentukan pemisahan yang terbaik. Sederhananya, Information gain melakukan perhitungan seberapa besar informasi yang dapat diberikan suatu atribut atau fitur tentang suatu kelas. Adapun perhitungan information gain dapat dilakukan dengan persamaan berikut:

$$Information\ Gain = E_k - (P_m E_m)$$

# Telaah Literature

## Evaluation Matrix

Evaluation Matrix merupakan alat untuk mengukur atau mengevaluasi performance dari model machine learning yang dibangun untuk mengolah data[5]. manfaat dari menggunakan evaluation matrix adalah untuk dapat mengukur ketetapan ataupun kesuksesan suatu model yang telah dibuat. Pada projek ini pengukuran evaluation matrix menggunakan confusion matrix.

## Confusion Matrix

Confusion Matrix merupakan salah satu metode evaluation matrix yang digunakan untuk mengukur nilai akurasi, presisi, dan recall dari model machine learning yang dibuat [6]. Dalam Confusion Matrix terdapat 4 komponen yaitu True Positif (TP), True Negativ (TN), False Positif (FP), False Negatif (FN)

# Evaluasi Model

Berikut adalah rumus dari akurasi yang memprediksi model yang dibangun baik dikelas positif dan kelas negatif

Rumus precision menghitung serta memprediksi ratio positif benar dibandingkan terhadap total data positif benar.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

# METODOLOGI - SEMMA

SAMPLE

EXPLORE

MODIFY

MODEL

ASSES

SEMMA mengacu pada tahapan atau metodologi yang dapat digunakan untuk proyek data mining maupun machine learning. Kerangka kerja ini terbukti telah banyak digunakan oleh industri sampai saat ini. Adapun SEMMA merupakan singkatan dari proses metodologinya yang terdiri dari lima tahapan yaitu Sample (pengambilan sampel), Explore (eksplorasi), Modify (modifikasi), Model (pemodelan), dan Asses (penilaian)

## SAMPLE

Data diambil dari  
situs resmi IDX pada  
Github dan juga  
Kaggle

No.	Atribut	Deskripsi
1	Code	Kode nama saham perusahaan
2	Name	Nama Perusahaan
3	ListingDate	Tanggal saham perusahaan pertama kali go public
4	Shares	Jumlah saham yang beredar dan persentase saham yang dimiliki oleh investor publik
5	ListingBoard	Kelas-kelas terbaginya saham menjadi 3 kategori
6	MarketCap	Kapitalisasi pasar perusahaan yang dihitung dengan men- galikan harga saham dengan jumlah saham yang beredar
7	DailyLastUpdate	Tanggal terakhir saham perusahaan di perbarui
8	Operational Lifetime	Lamanya perusahaan tercatat di IDX
9	Avg. Value	Rata-rata perdagangan harian saham perusahaan
10	Avg. Frequency	Rata-rata aktivitas perdagangan saham perusahaan dalam jangka waktu tertentu

## EXPLORE

Dilakukan exploratory data analysis untuk mencari tahu lebih lanjut mengenai sifat-sifat dari dataset yang dimiliki.

### Visualisasi Boxplot

untuk melihat outlier yang terdapat pada setiap kolom pada dataset.

### Visualisasi Histogram

untuk melihat distribusi data

### Visualisasi Countplot

untuk melihat jumlah data untuk masing-masing kelas pada dataset

## MODIFY

1.  
Normalisasi menggunakan MinMaxScaler, untuk memastikan nilai skala dari atribut dataset tidak terlalu jauh.

2.  
Standarisasi menggunakan StandardScaler, untuk memastikan penyebaran data menjadi mendekat kepada normal distribution

Modifikasi terhadap dataset sesuai dengan hasil eksplorasi pada tahap sebelumnya

3.  
Mapping kepada kolom kelas untuk mengubah dari categorical menjadi numerical.

4.  
Penghapusan outlier menggunakan teknik z-score, yang dimana setelah menggunakan zscore data menjadi 711 dari 733.

5.  
Teknik hybrid oversampling menggunakan SMOTETomek, untuk membuat dataset lebih balance.

## MODEL

Menggunakan K-Nearest Neighbors dan Decision Tree. Data yang sudah dimodifikasi tadi akan dibagi menjadi 2 set yaitu training dan testing set. Kemudian setiap set akan memiliki X dan Y dimana X merupakan independent variable dan Y merupakan dependent variable. Untuk menentukan parameter yang akan digunakan, digunakan library Grid Search yang berfungsi dalam mencari kombinasi parameter terbaik.

## ASSES

Metode evaluation yang digunakan adalah classification report, dimana dapat dilihat precision, recall, f1-score, dan support. Setelah dilihat classification report dari masing-masing model, kemudian akan dilakukan visualisasi confusion matrix.

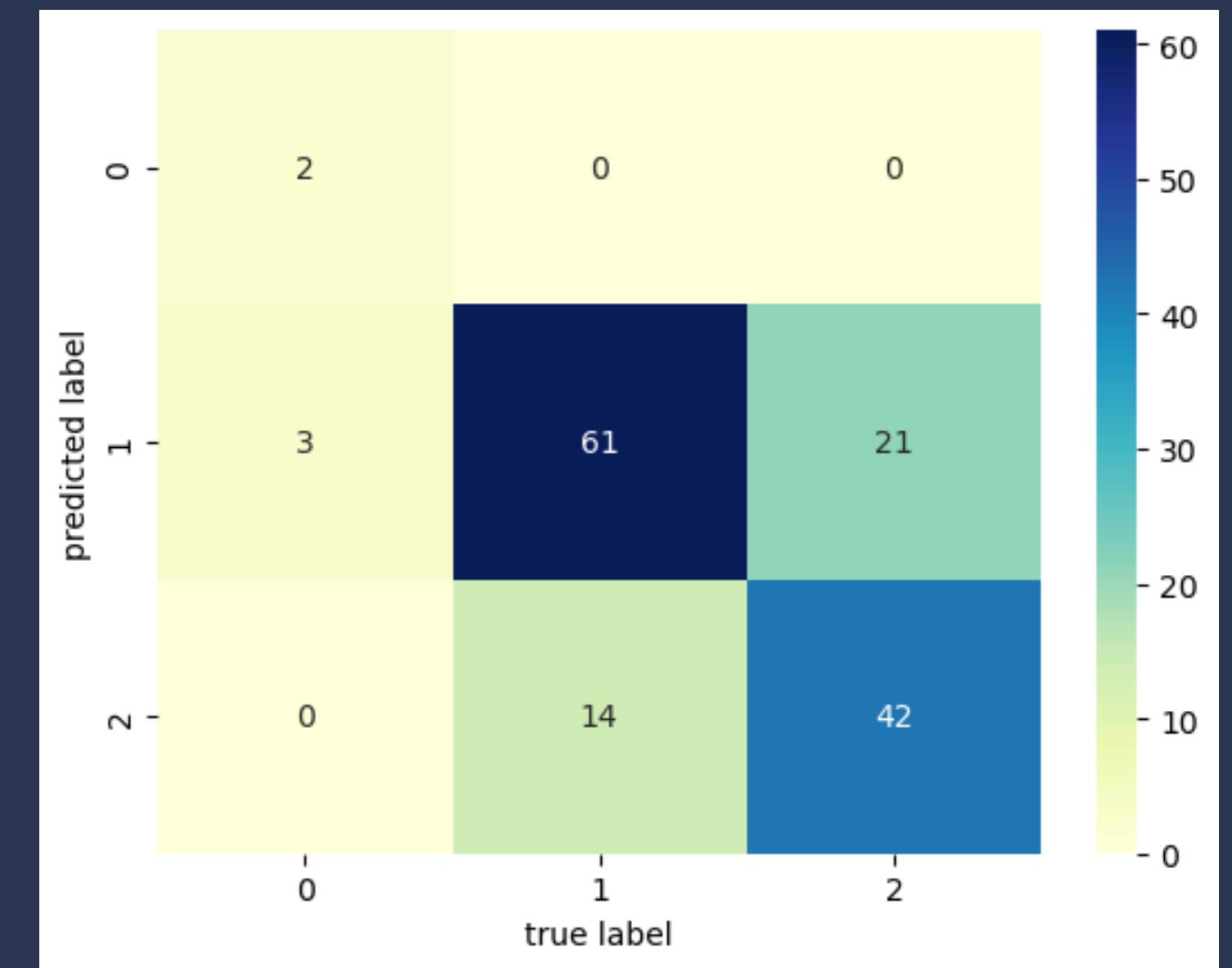
---

# HASIL DAN DISKUSI

## KNN (K-NEAREST NEIGHBOOR)

	Precision	Recall	F1-Score	Support
0	0.40	1.00	0.57	2
1	0.81	0.72	0.76	85
2	0.67	0.75	0.71	56
accuracy			0.73	143
macro avg	0.63	0.82	0.68	143
weighted avg	0.75	0.73	0.74	143

Evaluation KNN



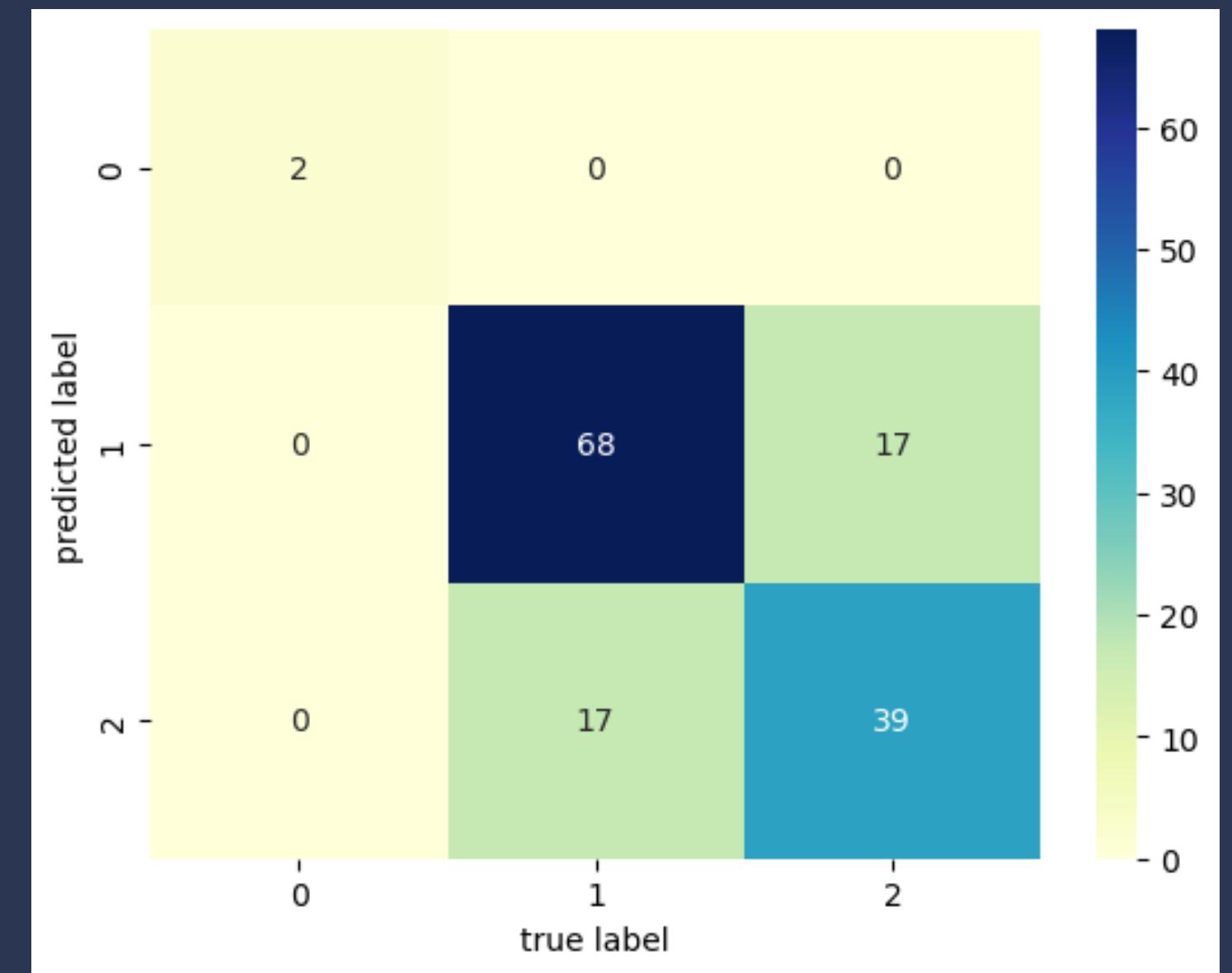
Confusion Matrix KNN

# HASIL DAN DISKUSI

## DECISION TREE

	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	2
1	0.879	0.80	0.80	85
2	0.70	0.70	0.70	56
accuracy			0.76	143
macro avg	0.83	0.83	0.83	143
weighted avg	0.76	0.76	0.76	143

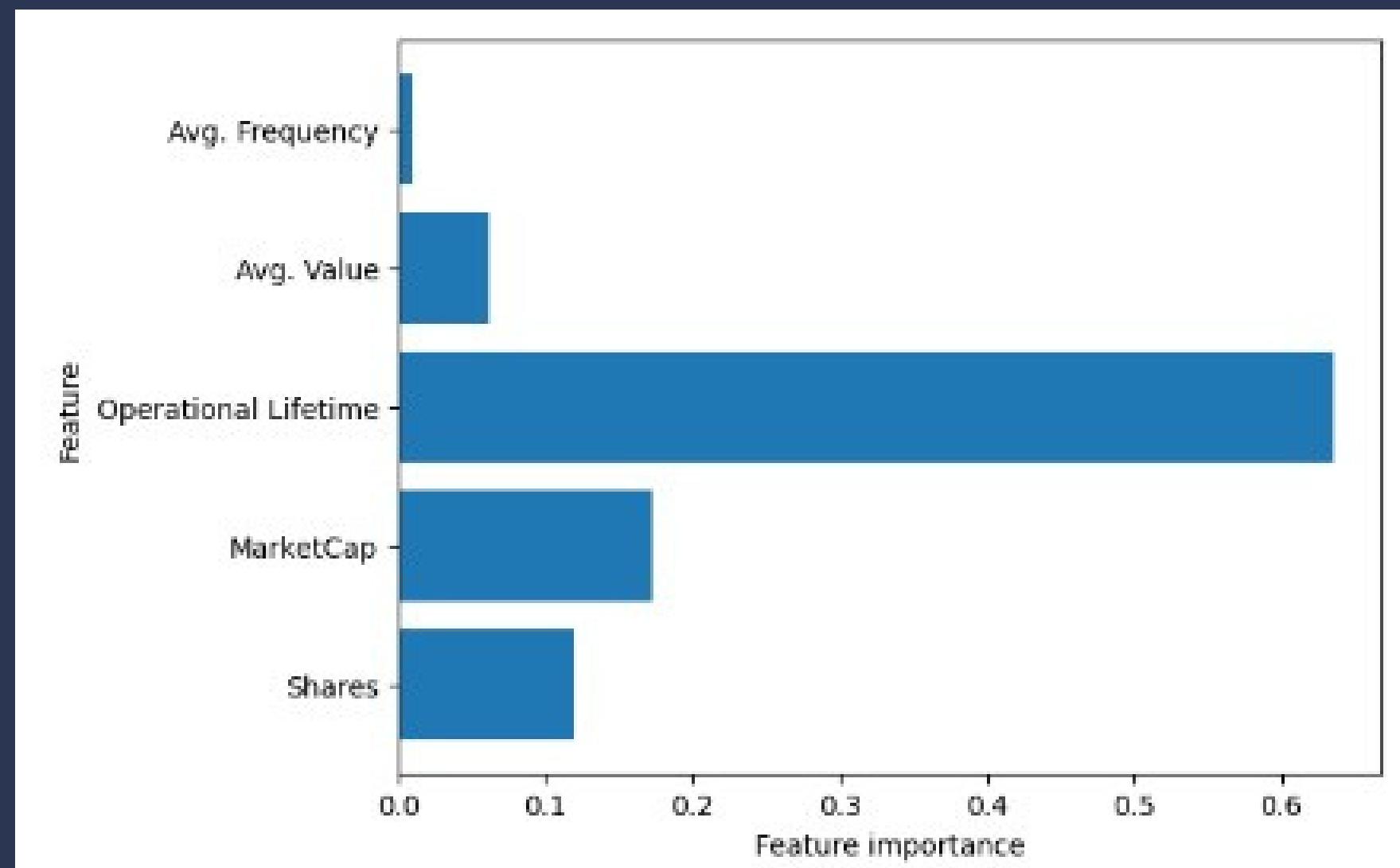
Evaluation Decision Tree



Confusion Matrix Decision Tree

# HASIL DAN DISKUSI

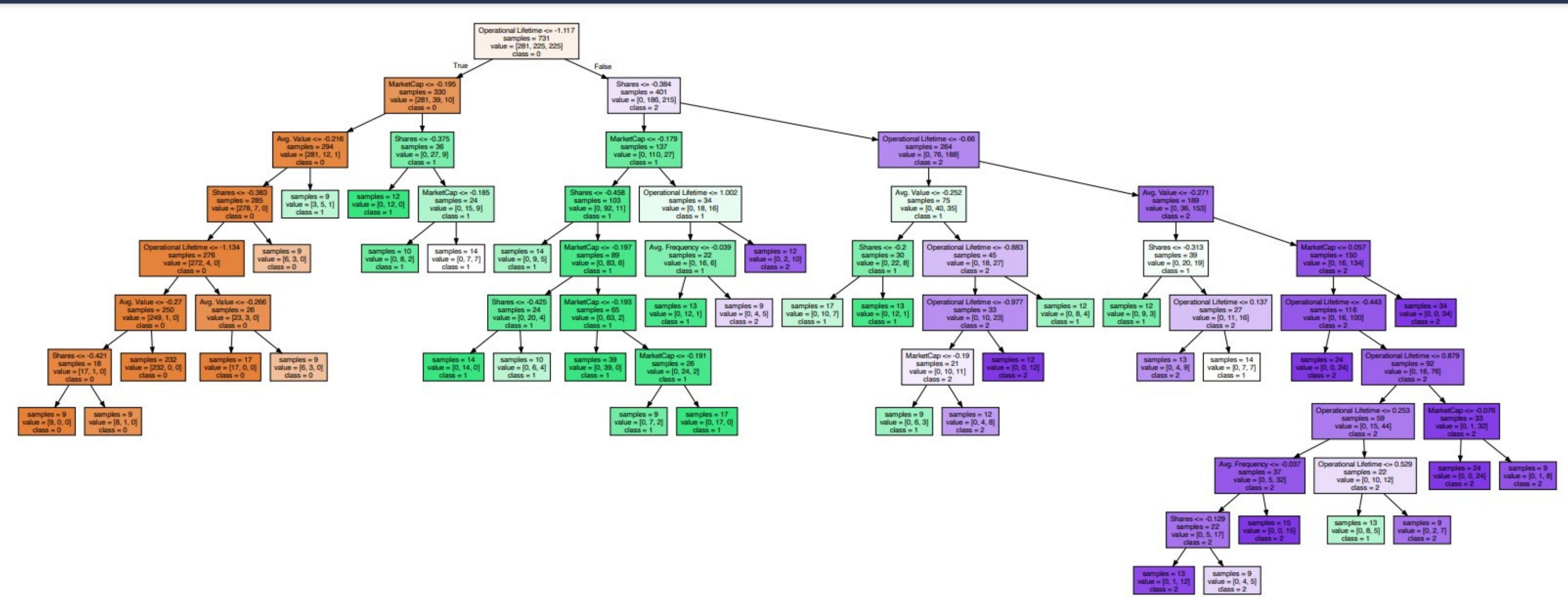
## DECISION TREE



Feature Importance Decision Tree

# HASIL DAN DISKUSI

# DECISION TREE



# Visualisasi Decision Tree

# KESIMPULAN

Sebagai kesimpulan dari penelitian ini, dari kedua model yang sudah dibuat dan dibandingkan hasil dan performanya, berikut merupakan beberapa poin yang bisa diambil untuk menjawab rumusan masalah yang sudah ditetapkan:

1. Faktor yang paling berpengaruh dalam melakukan klasifikasi saham kepada ketiga listing board yang tersedia adalah Operational Lifetime.
2. 5 Faktor yang digunakan dalam penelitian ini kurang dapat untuk mengklasifikasikan saham kepada listing board.
3. Algoritma yang memiliki hasil dan performa yang lebih baik adalah algoritma Decision Tree.



# REFERENSI

- D. SHAH, H. ISAH, AND F. ZULKERNINE, "STOCK MARKET ANALYSIS: A REVIEW AND TAXONOMY OF PREDICTION TECHNIQUES," VOL. 7(2), 2019.
- FIRAS, OMARI. "A COMBINATION OF SEMMA & CRISP-DM MODELS FOR EFFECTIVELY HANDLING BIG DATA USING FORMAL CONCEPT ANALYSIS BASED KNOWLEDGE DISCOVERY: A DATA MINING APPROACH." WORLD JOURNAL OF ADVANCED ENGINEERING TECHNOLOGY AND SCIENCES, VOL. 8, NO. 1, 2023.
- N. HOTZ, "WHAT IS SEMMA?," DATA SCIENCE PROCESS ALLIANCE, [HTTPS://WWW.DATASCIENCE-PM.COM/SEmma/](https://www.datascience-pm.com/semma/), 2023
- D. SREE AND S. BINDU, DATA ANALYTICS: WHY DATA NORMALIZATION (2018), [HTTPS://WWW.ACADEMIA.EDU/DOWNLOAD/74142822/9582.PDF](https://www.academia.edu/download/74142822/9582.PDF).
- V. R. PRASETYO, M. MERCIFIA, A. AVERINA, L. SUNYOTO, AND BUDIARJO, "JURNAL ILMIAH NERO VOL. 7 NO. 1 - REPOSITORY.UBAYA.AC.ID," PREDIKSI RATING FILM PADA WEBSITE IMDB MENGGUNAKAN METODE NEURAL NETWORK (2022), [HTTP://REPOSITORY.UBAYA.AC.ID/41805/1/268-890-1-PB.PDF](http://repository.ubaya.ac.id/41805/1/268-890-1-PB.PDF).
- X. GONG, F. ZHANG, T. LU, AND W. YOU, "COMPARATIVE ANALYSIS OF THREE OUTLIER DETECTION METHODS IN UNIVARIATE DATA SETS," 2022 3RD INTERNATIONAL CONFERENCE ON ELECTRONIC COMMUNICATION AND ARTIFICIAL INTELLIGENCE (IWECAI), PP. 209-213, 2022. DOI:10.1109/IWECAI55315.2022.00048
- M. KHUSHI ET AL., "A COMPARATIVE PERFORMANCE ANALYSIS OF DATA RESAMPLING METHODS ON IMBALANCE MEDICAL DATA," IEEE ACCESS, VOL. 9, PP. 109960–109975, 2021. DOI:10.1109/ACCESS.2021.3102399
- Z. WANG, C. WU, K. ZHENG, X. NIU, AND X. WANG, "SMOTETOMEK-BASED RESAMPLING FOR PERSONALITY RECOGNITION," IEEE ACCESS, VOL. 7, PP. 129678–129689, 2019. DOI:10.1109/ACCESS.2019.2940061
- E. ETRIYANTI, "PERBANDINGAN TINGKAT AKURASI METODE KNN DAN DECISION TREE DALAM MEMPREDIKSI LAMA STUDI MAHASISWA," JURNAL ILMIAH BINARY STMIK BINA NUSANTARA JAYA, VOL. 3, NO. 1, 2021.
- P. R. SIHOMBING AND A. M. ARSANI, "COMPARISON OF MACHINE LEARNING METHODS IN CLASSIFYING POVERTY IN INDONESIA IN 2018," JURNAL TEKNIK INFORMATIKA, VOL. 2(1), PP. 51–56, 6 2021.
- E. R. TAURAN, "PREDIKSI HARGA SAHAM PT BANK CENTRAL ASIA TBK BERDASARKAN DATA DARI BURSA EFEK INDONESIA MENGGUNAKAN METODE K-NEAREST NEIGHBORS (KNN),"TEIKA, VOL. 11, NO. 2, PP. 123–129, 2021.

# REFERENSI

- S. H. S. ROBIANTO AND U. RISTIAN, "PENERAPAN METODE DECISION TREE UNTUK MENGKLASIFIKASIKAN MUTU BUAH JERUK BERDASARKAN FITUR WARNA DAN UKURAN," JURNAL KOMPUTER DAN APLIKASI, VOL. 9(1), PP. 76–86, 2021.
- M. NABIPOUR, P. NAYYERI, H. JABANI, S. SHAHAB, AND A. MOSAVI, "PREDICTING STOCK MARKET TRENDS USING MACHINE LEARNING AND DEEP LEARNING ALGORITHMS VIA CONTINUOUS AND BINARY DATA; A COMPARATIVE ANALYSIS," IEEE ACCESS, VOL. 8, PP. 150 199–150 212, 2020.
- A. ARDIYANSYAH, P. A. RAHAYUNINGSIH, AND R. MAULANA, "ANALISIS PERBANDINGAN ALGORITMA KLASIFIKASI DATA MINING UNTUK DATASET BLOGGER DENGAN RAPID MINER," JURNAL KHATULISTIWA INFORMATIKA, VOL. 6, NO. 1, 6 2018.
- A. T. SARAGIH, D. AFRIDA, AND R. F. DINARA, "KLASIFIKASI JUMLAH KENDARAAN DI SUMATERA UTARA DAN SUMATERA BARAT MENGGUNAKAN ALGORITMA NAIVE BAYES," JCOM (JOURNAL OF COMPUTER), VOL. 3, NO. 1, PP. 11–16, 2023.
- R.AINUN, "PENILAIAN KELAYAKAN PINJAMAN NASABAH MENGGUNAKAN LOGIKA FUZZY  
• METODE MAMDANI," PH.D. DISSERTATION, INSTITUT TEKNOLOGI KALIMANTAN, 2021.
- R. RINDIYANI, A. PRIMADEWI, M. MAIMUNAH, AND A. H. PURWANTINI, "KLASIFIKASI PENJUALAN BERDASARKAN PLATFORM PADA UMKM OMAH BRANDED MENGGUNAKAN RANDOM FOREST," JURIKOM (JURNAL RISET KOMPUTER), VOL. 9, NO. 5, PP.,1520–1529, 2022.
- Y. SURYANA AND T. W. SEN, "THE PREDICTION OF GOLD PRICE MOVEMENT BY COMPARING NAIVE BAYES, SUPPORT VECTOR MACHINE, AND K-NN," JURNAL INFORMATIKA DAN SAINS, VOL. 4, NO. 2, PP. 112–120, 2021.

# THANKYOU!!

