

Tugas Minggu 11
SK5004 Pembelajaran Mesin Dan Kecerdasan Buatan
Senin, 5 Mei 2025

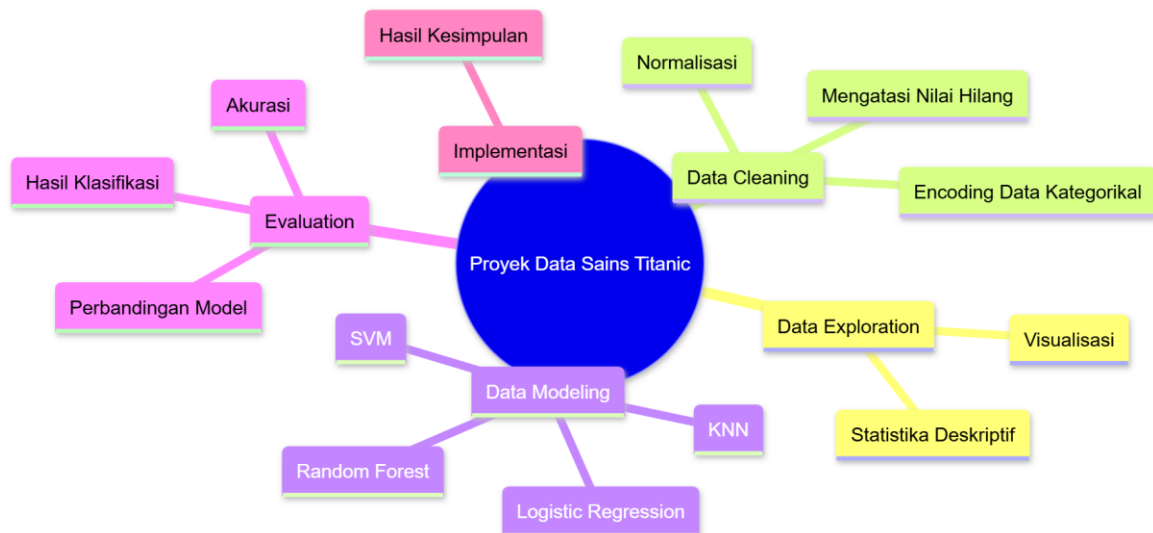
Nama : Naila Ratu Dianti

NIM : 20123019

MIND MAP DAN CONCEPT MAP

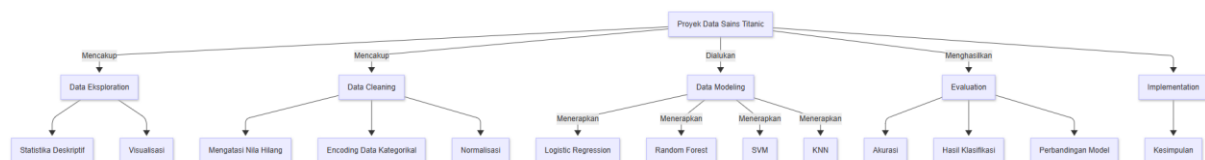
1. Mind Map

Mind map ini menunjukkan ide awal penelitian secara keseluruhan, dimulai dari tahap eksplorasi hingga evaluasi.



2. Concept Map

Concept map digunakan untuk menggambarkan penyusunan pertanyaan penelitian.



DATA SCIENCE AND PYTHON VENV

1. Konsep data science

Data science atau sering disebut juga Sains Data merupakan suatu cabang ilmu yang berfokus pada pengumpulan, pengolahan, analisis dan interpretasi data untuk mengidentifikasi pola,

tren, dan hasil yang berharga. Tujuan utamanya adalah menangkap informasi yang dapat digunakan untuk pengambilan keputusan dan pemecahan masalah dalam berbagai konteks.

2. Perbedaan data, data science, data scientist

Data

Data merupakan kumpulan informasi yang digambarkan dengan angka, simbol, gambar, atau uraian yang mempunyai arti pada suatu konteks tertentu. Data merepresentasikan suatu objek atau suatu kejadian. Adapun jenis data berdasarkan cara mendapatkannya, yaitu data primer dan data sekunder. Data primer merupakan data yang dikumpulkan secara langsung dari objek pengamatan. Data sekunder merupakan data yang didapatkan dari sumber lain yang telah ada.

Data Science

Data science merupakan suatu cabang ilmu yang menggabungkan berbagai bidang, seperti matematika, statistika, ilmu komputer, dan lainnya. Data science merupakan proses penggalian wawasan dari data yang melibatkan pengumpulan data, pembersihan data, analisis data. Melalui penggunaan teknik statistika, pemodelan data, dan kecerdasan buatan, data science membantu dalam memahami tren pasar, memperbaiki efisiensi operasional, dan mengoptimalkan pengambilan keputusan. Selain itu, data science juga dapat digunakan untuk berbagai masalah kompleks, seperti prediksi, analisis risiko, dan lain sebagainya.

Data Scientist

Data scientist merupakan seorang profesional yang memiliki keahlian dalam mengumpulkan, mengelola, menganalisis, dan menafsirkan data untuk menghasilkan pengetahuan yang berharga dan mendukung pengambilan keputusan. Tugas utama seorang data scientist meliputi pengumpulan data, pembersihan data, analisis data, pemodelan data, dan kemampuan komunikasi. Data scientist biasanya memiliki latar belakang dalam matematika, statistika, ilmu komputer, atau bidang terkait, dan bekerja di berbagai bidang seperti, teknologi, keuangan, kesehatan, dan lainnya.

3. Empat aspek penting dalam data science

Matematika, fondasi ini mencakup konsep dasar seperti fungsi, relasi, asumsi, kesimpulan, dan lainnya. Konsep-konsep ini membantu mendefinisikan dan memahami berbagai aspek. Dengan dasar matematika yang kuat, seorang data scientist dapat merancang solusi yang logis dan terstruktur untuk menyelesaikan masalah berbasis data.

Teknologi, fondasi ini berkaitan dengan penguasaan alat dan keterampilan pemrograman yang dibutuhkan dalam praktik data science. Pengetahuan dalam bahasa pemrograman seperti Python sangat penting.

Visualisasi, fondasi ini menekankan pada kemampuan menyampaikan informasi data melalui visualisasi yang tepat. Jenis grafik yang digunakan harus sesuai dengan tipe data dan tujuan komunikasi. Visualisasi juga bisa dibuat interaktif, misalnya dalam bentuk dashboard, untuk membantu pengambilan keputusan bisnis secara real-time.

Komunikasi, fondasi ini mencakup cara menyampaikan hasil analisis secara jelas dan efektif. Komunikasi dalam data science meliputi, menulis komentar dalam kode, membuat dokumentasi kode, menuliskan interpretasi hasil, Menyusun laporan teknis yang mudah dipahami.

4. Sumber PyPi untuk instal Jupyter Notebook, Matplotlib, Numpy

Link Jupyter Notebook : <https://pypi.org/project/notebook/>

Link Matplotlib : <https://pypi.org/project/matplotlib/>

Link Numpy : <https://pypi.org/project/numpy/>

5. Membuat virtual environment, instal packages, simpan informasi dalam requirements.txt, buat virtual environment dan gunakan requirement.txt

Membuat virtual environment 1

```
C:\Users\ASUS\belajar_python>python -m venv env1  
C:\Users\ASUS\belajar_python>env1\Scripts\activate
```

Instal packages

```
(env1) C:\Users\ASUS\belajar_python>pip install numpy pandas matplotlib
Collecting numpy
  Using cached numpy-2.2.5-cp313-cp313-win_amd64.whl.metadata (60 kB)
Collecting pandas
  Using cached pandas-2.2.3-cp313-cp313-win_amd64.whl.metadata (19 kB)
Collecting matplotlib
  Downloading matplotlib-3.10.3-cp313-cp313-win_amd64.whl.metadata (11 kB)
Collecting python-dateutil>=2.8.2 (from pandas)
  Using cached python_dateutil-2.9.0.post0-py2.py3-none-any.whl.metadata (8.4 kB)
Collecting pytz>=2020.1 (from pandas)
  Downloading pytz-2025.2-py2.py3-none-any.whl.metadata (22 kB)
Collecting tzdata>=2022.7 (from pandas)
  Downloading tzdata-2025.2-py2.py3-none-any.whl.metadata (1.4 kB)
Collecting contourpy>=1.0.1 (from matplotlib)
  Downloading contourpy-1.3.2-cp313-cp313-win_amd64.whl.metadata (5.5 kB)
Collecting cycler>=0.10 (from matplotlib)
  Using cached cycler-0.12.1-py3-none-any.whl.metadata (3.8 kB)
Collecting fonttools>=4.22.0 (from matplotlib)
  Downloading fonttools-4.58.0-cp313-cp313-win_amd64.whl.metadata (106 kB)
Collecting kiwisolver>=1.3.1 (from matplotlib)
  Downloading kiwisolver-1.4.8-cp313-cp313-win_amd64.whl.metadata (6.3 kB)
Collecting packaging>=20.0 (from matplotlib)
  Downloading packaging-25.0-py3-none-any.whl.metadata (3.3 kB)
Collecting pillow>=8 (from matplotlib)
  Downloading pillow-11.2.1-cp313-cp313-win_amd64.whl.metadata (9.1 kB)
Collecting pyparsing>=2.3.1 (from matplotlib)
  Downloading pyparsing-3.2.3-py3-none-any.whl.metadata (5.0 kB)
Collecting six>=1.5 (from python-dateutil>=2.8.2->pandas)
  Downloading six-1.17.0-py2.py3-none-any.whl.metadata (1.7 kB)
Using cached numpy-2.2.5-cp313-cp313-win_amd64.whl (12.6 MB)
Using cached pandas-2.2.3-cp313-cp313-win_amd64.whl (11.5 MB)
```

```
Downloading matplotlib-3.10.3-cp313-cp313-win_amd64.whl (8.1 MB)
  8.1/8.1 MB 2.5 MB/s eta 0:00:00
Downloading contourpy-1.3.2-cp313-cp313-win_amd64.whl (223 kB)
Using cached cycler-0.12.1-py3-none-any.whl (8.3 kB)
Downloading fonttools-4.58.0-cp313-cp313-win_amd64.whl (2.2 MB)
  2.2/2.2 MB 2.8 MB/s eta 0:00:00
Downloading kiwisolver-1.4.8-cp313-cp313-win_amd64.whl (71 kB)
Downloading packaging-25.0-py3-none-any.whl (66 kB)
Downloading pillow-11.2.1-cp313-cp313-win_amd64.whl (2.7 MB)
  2.7/2.7 MB 2.6 MB/s eta 0:00:00
Downloading pyparsing-3.2.3-py3-none-any.whl (111 kB)
Using cached python_dateutil-2.9.0.post0-py2.py3-none-any.whl (229 kB)
Downloading pytz-2025.2-py2.py3-none-any.whl (509 kB)
Downloading tzdata-2025.2-py2.py3-none-any.whl (347 kB)
Downloading six-1.17.0-py2.py3-none-any.whl (11 kB)
Installing collected packages: pytz, tzdata, six, pyparsing, pillow, packaging, numpy, kiwisolver, fonttools, cycler, python-dateutil, contourpy, pandas, matplotlib
Successfully installed contourpy-1.3.2 cycler-0.12.1 fonttools-4.58.0 kiwisolver-1.4.8 matplotlib-3.10.3 numpy-2.2.5 packaging-25.0 pandas-2.2.3 pillow-11.2.1 pyparsing-3.2.3 python-dateutil-2.9.0.post0 pytz-2025.2 six-1.17.0 tzdata-2025.2

[notice] A new release of pip is available: 24.2 -> 25.1.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

Menyimpan informasi pada requirement.txt

```
(env1) C:\Users\ASUS\belajar_python>pip freeze > requirements.txt
```

Membuat virtual environment 2

```
(env1) C:\Users\ASUS\belajar_python>python -m venv env2
```

```
(env1) C:\Users\ASUS\belajar_python>env2\Scripts\activate
```

Penggunaan requirement.txt

```

(env2) C:\Users\ASUS\belajar_python>pip install -r requirements.txt
Collecting contourpy==1.3.2 (from -r requirements.txt (line 1))
  Using cached contourpy-1.3.2-cp313-cp313-win_amd64.whl.metadata (5.5 kB)
Collecting cycler==0.12.1 (from -r requirements.txt (line 2))
  Using cached cycler-0.12.1-py3-none-any.whl.metadata (3.8 kB)
Collecting fonttools==4.58.0 (from -r requirements.txt (line 3))
  Using cached fonttools-4.58.0-cp313-cp313-win_amd64.whl.metadata (106 kB)
Collecting kiwisolver==1.4.8 (from -r requirements.txt (line 4))
  Using cached kiwisolver-1.4.8-cp313-cp313-win_amd64.whl.metadata (6.3 kB)
Collecting matplotlib==3.10.3 (from -r requirements.txt (line 5))
  Using cached matplotlib-3.10.3-cp313-cp313-win_amd64.whl.metadata (11 kB)
Collecting numpy==2.2.5 (from -r requirements.txt (line 6))
  Using cached numpy-2.2.5-cp313-cp313-win_amd64.whl.metadata (60 kB)
Collecting packaging==25.0 (from -r requirements.txt (line 7))
  Using cached packaging-25.0-py3-none-any.whl.metadata (3.3 kB)
Collecting pandas==2.2.3 (from -r requirements.txt (line 8))
  Using cached pandas-2.2.3-cp313-cp313-win_amd64.whl.metadata (19 kB)
Collecting pillow==11.2.1 (from -r requirements.txt (line 9))
  Using cached pillow-11.2.1-cp313-cp313-win_amd64.whl.metadata (9.1 kB)
Collecting pyparsing==3.2.3 (from -r requirements.txt (line 10))
  Using cached pyparsing-3.2.3-py3-none-any.whl.metadata (5.0 kB)
Collecting python-dateutil==2.9.0.post0 (from -r requirements.txt (line 11))
  Using cached python_dateutil-2.9.0.post0-py2.py3-none-any.whl.metadata (8.4 kB)
Collecting pytz==2025.2 (from -r requirements.txt (line 12))
  Using cached pytz-2025.2-py2.py3-none-any.whl.metadata (22 kB)
Collecting six==1.17.0 (from -r requirements.txt (line 13))
  Using cached six-1.17.0-py2.py3-none-any.whl.metadata (1.7 kB)
Collecting tzdata==2025.2 (from -r requirements.txt (line 14))
  Using cached tzdata-2025.2-py2.py3-none-any.whl.metadata (1.4 kB)
Using cached contourpy-1.3.2-cp313-cp313-win_amd64.whl (223 kB)
Using cached cycler-0.12.1-py3-none-any.whl (8.3 kB)
Using cached fonttools-4.58.0-cp313-cp313-win_amd64.whl (2.2 MB)

Using cached kiwisolver-1.4.8-cp313-cp313-win_amd64.whl (71 kB)
Using cached matplotlib-3.10.3-cp313-cp313-win_amd64.whl (8.1 MB)
Using cached numpy-2.2.5-cp313-cp313-win_amd64.whl (12.6 MB)
Using cached packaging-25.0-py3-none-any.whl (66 kB)
Using cached pandas-2.2.3-cp313-cp313-win_amd64.whl (11.5 MB)
Using cached pillow-11.2.1-cp313-cp313-win_amd64.whl (2.7 MB)
Using cached pyparsing-3.2.3-py3-none-any.whl (111 kB)
Using cached python_dateutil-2.9.0.post0-py2.py3-none-any.whl (229 kB)
Using cached pytz-2025.2-py2.py3-none-any.whl (509 kB)
Using cached six-1.17.0-py2.py3-none-any.whl (11 kB)
Using cached tzdata-2025.2-py2.py3-none-any.whl (347 kB)
Installing collected packages: pytz, tzdata, six, pyparsing, pillow, packaging, numpy, kiwisolver, fonttools, cycler, python-dateutil, contourpy, pandas, matplotlib
Successfully installed contourpy-1.3.2 cycler-0.12.1 fonttools-4.58.0 kiwisolver-1.4.8 matplotlib-3.10.3 numpy-2.2.5 packaging-25.0 p
andas-2.2.3 pillow-11.2.1 pyparsing-3.2.3 python-dateutil-2.9.0.post0 pytz-2025.2 six-1.17.0 tzdata-2025.2

[notice] A new release of pip is available: 24.2 -> 25.1.1
[notice] To update, run: python.exe -m pip install --upgrade pip

```

PYTHON UNTUK MACHINE LEARNING

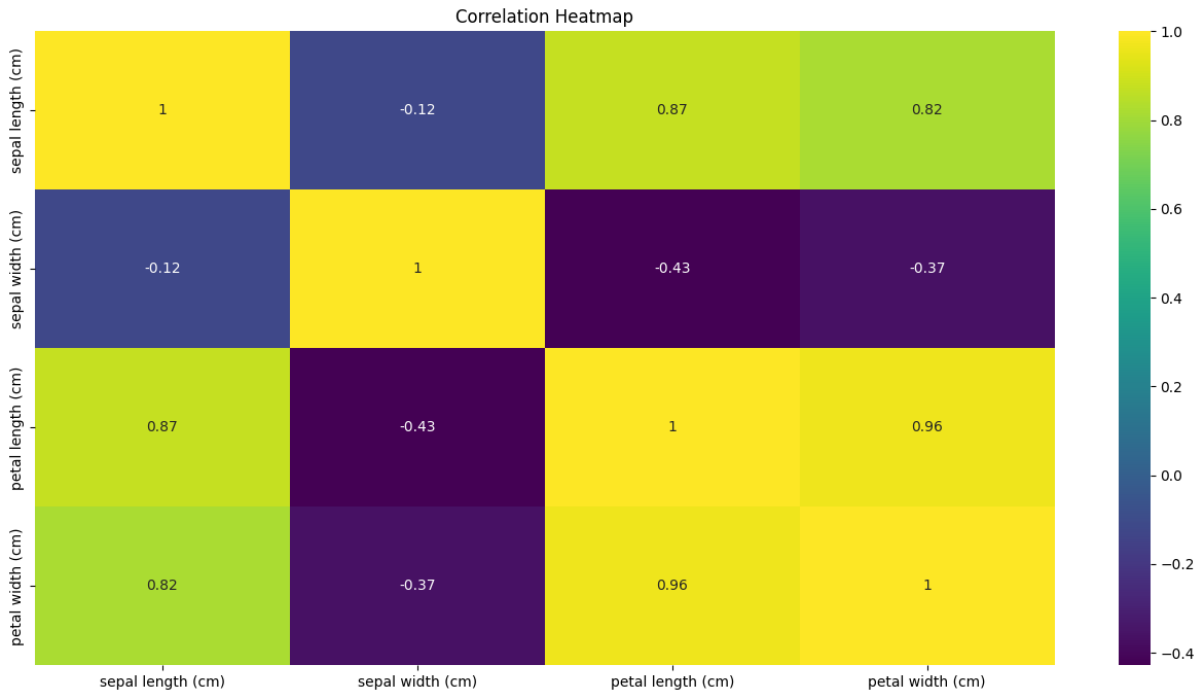
Eksplorasi data bertujuan untuk memahami karakteristik umum dari dataset sebelum dilakukan proses data modeling. Pada kasus ini, eksplorasi dikaukan terhadap dataset Iris yang terdiri dari 150 sampel bunga Iris yang diklasifikasikan dalam tiga spesies, yaitu Iris-setosa, Iris-versicolor, dan Iris-virginica. Berikut sari numerik dari masing-masing fitur.

Tabel 1 Sari Numerik Fitur

Fitur	Mean	Standar Deviasi	Minimum	Maximum
Sepal length	5,843	0,828	4,300	7,900
Sepal width	3,507	0,436	2,000	4,400
Petal length	3,758	1,765	1,000	6,900
Petal width	1,199	0,762	0,100	2,500

Berdasarkan tabel 1 diperoleh bahwa petal length dan petal width memiliki rentang nilai yang paling lebar dan memiliki variansi yang besar. Sedangkan sepal length dan sepal width memiliki variansi yang lebih kecil.

Selanjutnya untuk memahami hubungan antar fitur, dilakukan visualisasi korelasi. Korelasi dihitung berdasarkan nilai Pearson, yaitu nilai antara -1 hingga 1 yang menunjukkan kekuatan dan arah hubungan linear dalam fitur. Berikut hasil korelasi yang diperoleh.



Gambar 1 Heatmap Korelasi Fitur

Hasil visualisasi menunjukkan bahwa petal length dan petal width memiliki korelasi positif yang sangat kuat, yaitu 0,96. Sedangkann sepal width dan sepal sepal length memiliki korelasi negatif, yaitu -0,12. Hubungan antara sepal length dengan petal memiliki hubungan yang kuat, yaitu 0,87 dan 0,82. Sedangkan sepal width dengan petal memiliki hubungan yang negatif lemah.

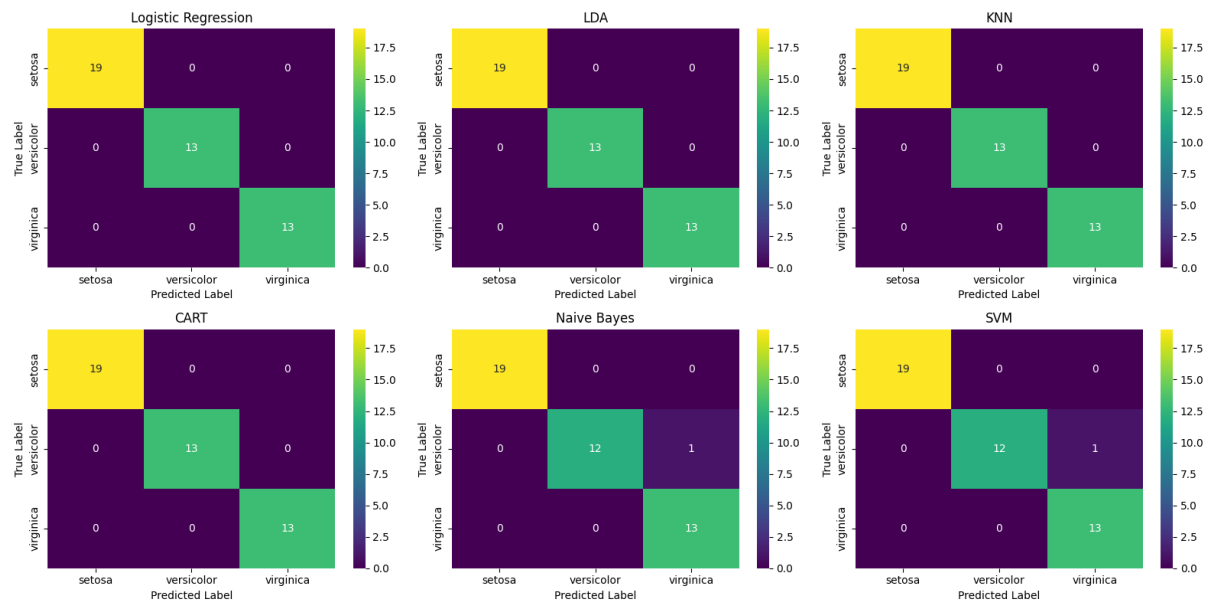
Setelah dilakukan training pada keenam model klasifikasi, selanjutnya dilakukan proses evaluasi performa model. Evaluasi dilakukan dengan melihat nilai Accuracy yaitu persentase prediksi yang benar terhadap seluruh data uji. Precision yaitu ketepatan model dalam memprediksi setiap kelas. Recall yaitu kemampuan model dalam menangkap semua data dari tiap kelas. F1-score yaitu rata-rata antara precision dan recall, mengukur keseimbangan keduanya. Berikut hasil evaluasi keenam model terhadap data uji.

Tabel 2 Hasil Evaluasi Model

Model	Accuracy	Setosa			Versicolor			Virginica		
		Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
LR	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
LDA	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
KNN	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Decision Tree	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Naïve Bayes	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
SVM	0,970	1,000	1,000	1,000	1,000	0,890	0,940	0,920	1,000	0,960

Model LR, LDA, KNN, Decision Tree, dan Naïve Bayes mencapai akurasi 100%, serta nilai precision, recall, dan F1-score yang juga sempurna di semua kelas. Hal ini menunjukkan bahwa model mampu mengklasifikasikan seluruh data uji dengan benar tanpa kesalahan. Model SVM menunjukkan sedikit penurunan performa dengan akurasi 97%. Hal ini menunjukkan bahwa model SVM sempat salah memprediksi beberapa sampel Versicolor sebagai Virginica.

Confusion matrix digunakan untuk mengevaluasi performa model klasifikasi secara detail. Angka dalam kotak menyatakan jumlah sampel. Idealnya, seluruh prediksi benar berada pada diagonal utama. Berikut visualisasi confusion matrix.



Gambar 2 Visualisasi Confusion Matrix

Setosa diklasifikasi 100% benar oleh semua model, hal ini menunjukkan bahwa kelas ini paling mudah dibedakan. Versicolor dan Virginica paling berpotensi terjadi misclassification. Model terbaik secara visualisasi adalah LR, LDA, KNN, dan Decision Tree. Model dengan sedikit error yaitu Naïve Bayes dan SVM.

Code access

https://colab.research.google.com/drive/1TDCJ9wzPoig_PlrqhGLt7ExHcq-Tu4ii?usp=sharing