

Implementasi Machine Learning menggunakan Python Library: Scikit-Learn (Supervised dan Unsupervised Learning)

Mohammad Naufal Fahmi

Sekolah Tinggi Agama Islam Nurul Islam Mojokerto

Email: naufal@nuris.ac.id

DOI: [10.52620/sainsdata.v1i2.31](https://doi.org/10.52620/sainsdata.v1i2.31)

ABSTRAK

Machine learning merupakan suatu pendekatan kecerdasan buatan yang mempunyai kegunaan dalam meniru sikap manusia dalam pengambilan keputusan berdasarkan pengalaman data yang dimiliki. Dataset yang digunakan dalam penelitian ini digunakan untuk proses inputan program yang berguna untuk menganalisis data target dengan benar. Dataset yang digunakan yaitu dataset iris yang diperoleh dari repository dataset publik oleh UCI Repository. Metodologi penelitian yang digunakan yaitu harus melalui empat tahapan yaitu: 1) tahap pemahaman masalah, yaitu masalah mengenai bunga iris yang memiliki macam warna dan memiliki panjang sepals dan petals untuk dikategorikan ke dalam spesies bunga iris sesuai dengan kecocokan data; 2) tahap pemilihan teknik, yaitu pemilihan teknik yang sesuai dengan tujuan dari pengelompokan bunga-bunga iris ke dalam spesiesnya iris sentosa, iris versicolor, atau iris virginica; 3) tahap persiapan data, yaitu menyiapkan dataset, kemudian menggunakan bahasa python dengan platform Jupiter Notebook atau Pycharm yang menyediakan beberapa library untuk membuat machine learning, dan; 4) tahap implementasi, yaitu merupakan tahap akhir dalam menyelesaikan permasalahan. Proses implementasi machine learning menggunakan teknik supervised learning antara lain digunakan pada jenis klasifikasi dengan menggunakan algoritma K-Neighbors Classifier dan pada jenis regresi menggunakan algoritma Linier Regression, sedangkan teknik unsupervised learning antara lain digunakan pada jenis reduksi dimensi dengan menggunakan algoritma PCA dan pada jenis klasterisasi/pengelompokan dengan menggunakan algoritma K-Means. Kedua teknik tersebut memanfaatkan library scikit-learn. Sehingga bahasa pemrograman Python yang diprogram memiliki keahlian yaitu memberikan hasil keluaran sesuai input dari dataset, metode dan algoritma yang digunakan.

Kata kunci: scikit-learn, supervised learning, unsupervised learning, machine learning, python.

Abstract: *Machine learning* is an artificial intelligence approach that has a uses in imitating human attitudes in making decisions based on the experience of the data. The dataset used in this study is used for the program input process which is useful for analyzing target data correctly. The dataset used is the iris dataset obtained from the public dataset repository by the UCI Repository. The research methodology used is that it has to go through four stages, namely: 1) the stage of understanding the problem, namely the problem of iris flowers that have different colors and lengths of sepals and petals to be categorized into iris species according to the suitability of the data; 2) the technique selection stage, namely the selection of a technique suitable for the purpose of grouping iris flowers into iris sentosa, iris versicolor, or iris virginica; 3) the data preparation stage, namely preparing the dataset, then using python language with the Jupiter Notebook platform or Pycharm, which provides several libraries for making machine learning, and; 4) the implementation stage, which is the final stage in solving the problem. The process of implementing machine learning using supervised learning techniques, among others, is used in the type of classification using the K-Neighbors Classifier algorithm and in the type of regression using the Linear Regression algorithm, while unsupervised learning techniques are used, among others, in the type of dimensional reduction using the PCA algorithm and in the type of clustering. grouping using the K-Means algorithm. Both techniques make use of the scikit-learn library. So that the Python programming language that is programmed has the expertise to provide output according to the input of the dataset, methods and algorithms used.

Keywords: scikit-learn, supervised learning, unsupervised learning, machine learning, python.

PENDAHULUAN

Machine Learning (ML) atau pembelajaran mesin ialah sebuah pendekatan dalam intelegensi artifisial (AI) yang memiliki kegunaan untuk meniru sikap manusia dalam hal pengambilan sebuah keputusan serta mengaplikasikannya secara otomatis (Ahmad, 2017). Sistem kerja dari *mechine learning* yaitu dengan cara mengerjakan suatu tugas (*task*) tanpa disertai pemrograman secara eksplisit dan proses tersebut didasari oleh sebuah pengalaman (*experience*). Sehingga fokus dari pengaplikasian pembelajaran mesin ini mengacu pada pengembangan sistem untuk menjadi pebelajar mandiri dalam memutuskan sesuatu, tanpa adanya pemrograman kembali, serta mampu beradaptasi terhadap terjadinya suatu perubahan. Sebagai input, semua kumpulan dari data yang besar (*big data*) dilakukan analisis dalam *machine learning* sampai menemukan suatu pola tertentu (Wahyono, 2018).

Pemrograman tradisional yang mana data dan program (*input*) kemudian dijalankan untuk mendapatkan keluaran (*output*), sangat berbeda dengan pemrograman *machine learning*, dimana data dan *output* digunakan sebagai masukan kemudian dijalankan pada komputer untuk membuat suatu program, dan program tersebut bisa digunakan untuk pemrograman tradisional (Brownlee, 2015). *Machine learning* memiliki beberapa teknik, namun secara garis besar *machine learning* ini mempunyai dua teknik dasar belajar, yaitu *supervised learning* (pembelajaran terarah) dan *unsupervised learning*.

Teknik *supervised learning* merupakan suatu metode yang dapat diimplementasikan pada *machine learning* yang dapat memperoleh informasi/prediksi dimana membutuhkan data/fitur pada setiap sampel dan label yang bersesuaian sebagai alat masukan (*input*). Jenis dari teknik *supervised learning* yaitu *classification* dan *regression*. *Classification* yang mana bentuk labelnya yaitu kategori sedangkan *regression* yang mana bentuk labelnya yaitu bilangan real. Tujuan dari teknik ini yaitu agar bisa memberikan target terhadap keluaran dengan cara membandingkan pada pengalaman belajar dari masa lalu. Sedangkan teknik *unsupervised learning* merupakan kebalikan dari *supervised learning* yaitu suatu metode yang dapat diimplementasikan pada pembelajaran mesin yang mana hanya membutuhkan data saja sebagai inputan. Jenis dari teknik ini yaitu *clustering* dan *demensionality reduction*. Cara kerja *clustering* yaitu dengan cara mengelompokkan bentuk data/*behavior* datanya berdasarkan kemiripan dari data sedangkan *demensionality reduction* bertujuan untuk membuat data menjadi data yang dapat divisualisasikan.

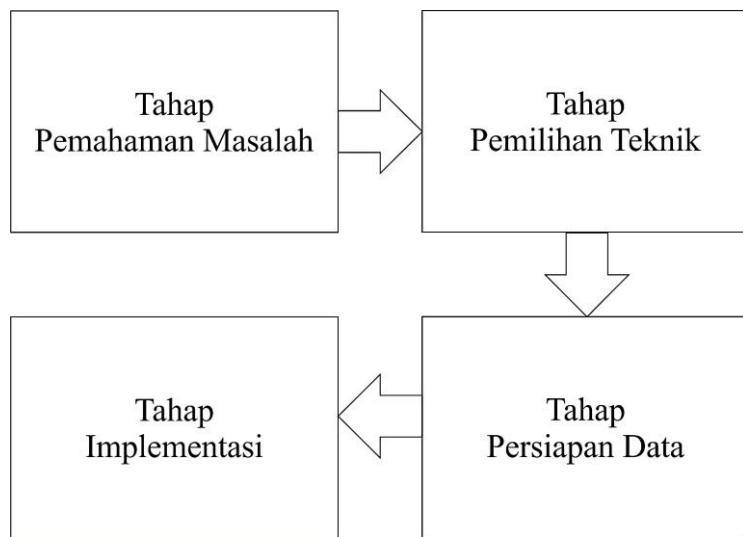
Dalam mengimplementasikan kedua teknik diatas dalam pemrograman python maka diperlukan *library Scikit-Learn*. *Scikit-Learn* adalah *library* Python yang dikembangkan untuk memberikan kemudahan dalam pengkodean *machine learning* dalam Python, melalui "API" yang dikembangkan oleh beberapa kontibutor dari berbagai negara yang mana biasanya digunakan pada dunia industri dan dunia akademisi (Pedregosa dan Varoquaux, Ga'el, 2011). Bisa dikatakan bahwa *library scikit-learn* ini sudah termasuk performa yang *optimized* karena *library* ini dirancang di atas modul *NumPy* (*Numerical Python*) dan *SciPy* (*Sientific Python*), sehingga perhitungan di dalamnya menjadi lebih efisien . Namun tedapat kelemahan pada *library* ini yaitu *scikit-learn* tidak dianjurkan untuk dipergunakan pada desain data yang lebih besar.

Pada penelitian ini, proses implementasi *machine learning* menggunakan bahasa pemrograman Python. Python merupakan bahasa perograman tingkat tinggi yang mana dalam segi perancangan bahasa berfokus pada tingkat keterbacaan dari *programming code* sehingga mudah untuk dipelajari. Python bisa dijalankan melalui berbagai layanan sistem operasi Windows (python), Linux, Android (Google Play), Mac OS dan lainnya (Arifin dan Muktyas, 2018). Python dikenal sebagai bahasa yang menggabungkan kapabilitas, kemampuan dengan sintaksis kode yang jelas dan dilengkapi dengan fungsionalitas pustaka standar yang besar. Oleh karena hal itu, jika suatu program yang dirancang dalam bahasa ini, maka program tersebut harus melalui proses terlebih dahulu sebelum dijalankan dalam komputer sehingga memerlukan waktu yang sedikit lama dibandingkan dengan bahasa pemrograman tingkat rendah.

Python progmaming diimplementasikan melalui intepreter interaktif yaitu *Jupyter Notebook* yang diinstall melalui sebuah perangkat lunak *stack* yang bernama *Anaconda Data Sceience Platform* (Wahyu, Tosida dan Andria, 2019) dan bisa juga diimplementasikan melalui Pycharm. *Anaconda* merupakan *distribution package* Python dari *Continuum Analytics* yang didalamnya berisi paket Python dengan berbagai paket tambahan, yang berguna untuk melakukan pemrograman data *science*, numerik, maupun platform yang *userfreindly*. *Anaconda Data Platform* (<http://repo.anaconda.com/>, no date) dapat didownload melalui laman resminya yaitu <https://www.anaconda.com/>. Kelebihan dari *Anaconda Data Platform* dapat mempermudah melaksanakan proses konfigurasi semua *library package* yang berhubungan dengan *modeling* yang sedang dikerjakan dan ada di dalam satu tempat. Python merupakan bahasa *default* dari *Anaconda*, sehingga Python yang digunakan dalam proses implementasi *machine learning* ini disarankan menggunakan Python versi 3.7 ke atas, karena python versi 2.x sudah berakhir masa pengembangannya pada tahun 2020.

METODE PENELITIAN

Dasar membuat sebuah *machine learning* memerlukan langkah-langkah yang ditunjukan pada gambar 2.1 (Retnoningsih dan Pramudita, 2020). Berikut merupakan penjelasan dari beberapa tahapan-tahapannya: 1) Tahap memahami permasalahan yang ingin diselesaikan. Mengetahui latar belakang masalah serta tujuan dari suatu pemecahan masalah yang diperoleh merupakan suatu keharusan sebagai dasar dan pedoman dalam pembuatan *machine learning*; 2) Tahap memilih teknik dan jenisnya dari *machine learning* yang digunakan. Mencocokkan antara permasalahan dengan teknik merupakan tahapan kedua yang harus dilaksanakan demi memperoleh output yang diinginkan; 3) Tahap menyiapkan data. Menyiapkan data secara lengkap yang terdiri dari variabel dan fitur-fitur di dalam data tersebut merupakan sebuah hal yang harus dilakukan, sehingga data tersebut dapat diolah dan dapat dijadikan acuan untuk melakukan tahap selanjutnya; 4) Tahap mengimplementasikan, melakukan pengkodean (*programming*) untuk menghasilkan output yang sesuai dengan tujuan dari permasalahan yang ada.



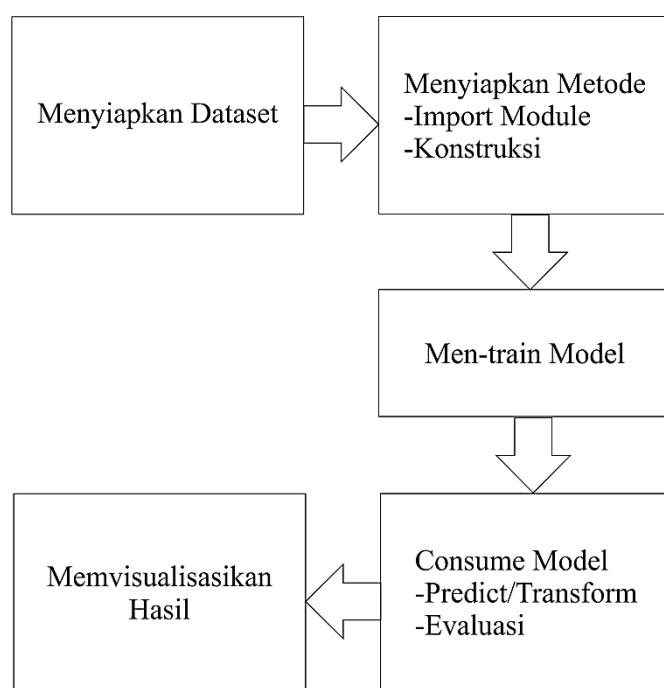
Gambar 2. 1 Tahapan-Tahapan Membuat Machine Learning
(Sumber: Wahyono (Retnoningsih: 2020))

Beberapa tahapan tersebut diimplementasikan guna membuat *machine learning* menggunakan bahasa python. Teknik yang digunakan dalam *supervised lerning* yaitu pengaplikasian klasifikasi (*classification*) dan regresi (*regression*) sedangkan teknik yang akan digunakan pada *unsupervised learning* yaitu pengelompokan (*clustering*) dan reduksi dimensi (*demensionality reduction*). Sedangkan dataset

yang digunakan dalam proses tahapan persiapan data yaitu dataset *Iris Tectorum* (<http://archive.ics.uci.edu/ml/datasets/Iris>, no date), yang merupakan dataset populer yang diperoleh dari *repository* dataset publik oleh *UCI Repository* <http://archive.ics.uci.edu/ml/index.php>. Dataset iris ini merupakan data yang mendeskripsikan tiga spesies bunga iris (kelas target) yaitu iris setosa, iris versicolor, dan iris virginica. Dataset iris ini memiliki fitur panjang sepal, lebar sepal, panjang petal, dan lebar petal dalam centimeter. Terdapat sekitar 150 data pada data set tersebut. Sehingga dengan dataset tersebut dapat pengimplementasian beberapa teknik dalam pembuatan *machine learning*, yang akan memudahkan dalam proses mendeteksi data baru yang menggunakan parameter dataset tersebut.

HASIL DAN PEMBAHASAN

Proses implementasi penelitian ini yaitu dalam pembuatan program *machine learning*, harus dilakukan sesuai dengan tahapan-tahapan yang sudah dipaparkan pada gambar 2.1. Tahap pertama yaitu pemahaman masalah, sehingga pada kasus ini dimulai dengan proses memahami permasalahan seperti bunga iris memiliki berbagai variasi panjang dari sepal dan petal, maupun warna bunganya, maka diperlukan tahapan kedua yaitu pemilihan metode/teknik yang akan digunakan. Pemilihan teknik dalam membuat *machine learning* menggunakan scikit-learn, pada kasus ini akan dipilih dua teknik yaitu *supervised learning* dan *unsupervised learning*. Dimana *supervised learning* akan menggunakan proses klasifikasi dan regresi sedangkan *unsupervised learning* menggunakan reduksi dimensi dan pengelompokan. Tahap ketiga yaitu persiapan data, pada kasus ini data diperoleh dari *repository public dataset* <http://archive.ics.uci.edu/ml/index.php>, yang berjudul dataset *Iris*. Kemudian data tersebut diunduh pada situs tersebut guna dijadikan bahan informasi dalam menyelesaikan tujuan penelitian. Tahapan terakhir yaitu proses implementasi, setelah proses memahami permasalahan, proses menentukan teknik yang digunakan dalam pembuatan *machine learning* dan persiapan dataset. Maka diperlukan langkah yaitu menyiapkan interface *Jupiter Notebook* atau Pycharm, kemudian mengimplementasikan langkah-langkah sebagaimana pada gambar 3.1 ini dalam pembuatan *machine learning* pada tiap-tiap tekniknya menggunakan *library scikit-learn* dengan bahasa pemrograman python.



Gambar 3. 1 Tahap-Tahap Implementasi Pembuatan Mechanie Learning
(Sumber: Hasil Penelitian 2021)

Teknik *Supervised Learning* menggunakan jenis Klasifikasi (*Classification*) (*K-Neighbors Classifier*)

Langkah pembuatan *machine learning* dengan teknik *supervised learning* jenis klasifikasi membutuhkan pemanggilan *library* *sklearn* yang *library*-nya digunakan untuk berbagai metode yang akan diterapkan dalam pengoperasian *machine learning*, dalam kasus ini *sklearn* memanggil dataset iris dan *library* *K-NeighborsClassifier*. Berikut ini merupakan *source code* yang *simple* dari penerapan beberapa tahapan dalam pembuatan *machine learning* menggunakan *K-Neighbors Classifier* yang dapat ditunjukkan melalui tabel 3.1.

Tabel 3.1. *Source Code* Tahap Implementasi *Machine Learning* menggunakan *K-Neighbors Classifier*

Keterangan	Source Code
Me-import library	<pre>from sklearn.datasets import load_iris from sklearn.neighbors import KNeighborsClassifier</pre>
Menyiapkan dataset	<pre>iris = load_iris() X, y = iris.data, iris.target</pre>
Menyiapkan metode klasifikasi	<pre>model = KNeighborsClassifier()</pre>
Me-train model	<pre>model.fit(X, y)</pre>
Me-consume model	<pre>new_data = [0,1,1,0] res = model.predict([new_data]) print(iris.target_names[res])</pre>
Hasil output	<pre>['setosa']</pre>

Dari langkah-langkah di atas, jika code diatas yang menggunakan algoritma *K-NeighborsClassifier* ketika di-*running* maka akan memprediksi dengan hasil *output* "setosa" yang berupa array.

Teknik *Supervised Learning* menggunakan jenis Regresi (*Regression*)

Pemrograman dengan teknik regresi ini dilanjutkan dengan pembuatan *cell* baru pada halaman *Jupyter Notebook*. Selanjutnya memanggil *library*-*library* yang dibutuhkan dalam pemrograman seperti: a) numpy, *library* yang digunakan untuk kebutuhan matematis dan *scientific*; b) matplotlib, *library* yang memiliki kegunaan untuk membuat grafik plot visualisasi, dan; c) sklearn, *library* yang berguna untuk membuat *machine learning* yang mana disini untuk proses pemanggilan *library* *Linier Regression*. Kode pemrograman regresi ini dapat dilihat pada tabel 3.2a berikut:

Tabel 3.2a. *Source Code* Tahap Implementasi *Machine Learning* menggunakan *Linier Regression*

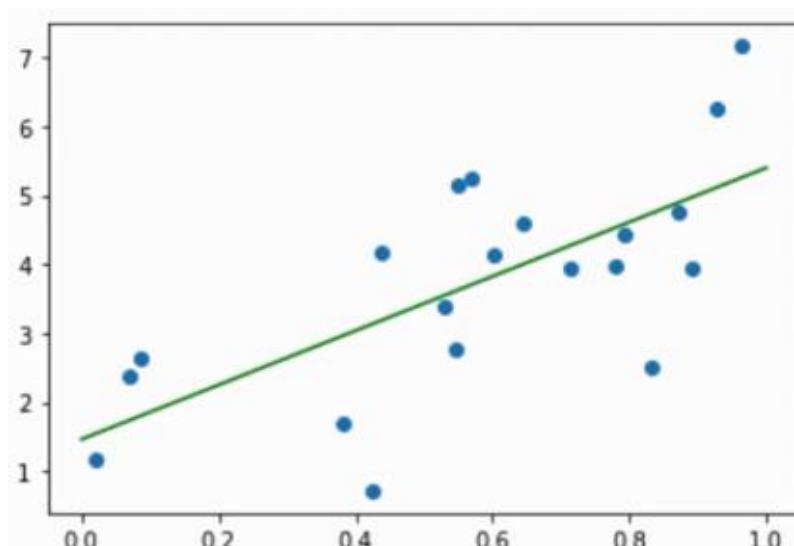
Keterangan	Source Code
Me-import library	<pre>import numpy as np import matplotlib.pyplot as plt from sklearn.linear_model import LinearRegression np.random.seed(0)</pre>
Menyiapkan dataset	<pre>X = np.random.random(size=(20, 1)) y = 3*X.squeeze() + 2 + np.random.randn(20)</pre>
Menyiapkan metode regresi	<pre>model = LinearRegression()</pre>
Me-train model	<pre>model.fit(X, y)</pre>
Me-consume model	<pre>X_fit = np.linspace(0,1,100)[:, np.newaxis] y_fit = model.predict(X_fit)</pre>

Dari langkah-langkah di atas, apabila menginginkan penampilan visualisasi dari hasil perograman di atas maka diperlukan langkah pada tabel 3.2b. berikut:

Tabel 3.2b. *Source Code* Tahap Visualisi Hasil menggunakan *Linier Regression*

Keterangan	Source Code
Memvisualisasikan hasil	<pre>plt.scatter(X,y); plt.plot(X_fit.squeeze(), y_fit, color='green');</pre>

Berdasarkan pengimplementasian *sorce code* pada tabel 3.2a dan 3.2b di atas, maka akan menghasilkan *output* visualisasi ketika di-*Run Cells*, yaitu akan menghasilkan tampilan sebagaimana pada gambar 3.3 berikut:



Gambar 3. 3 Hasil Visualisasi Regresi Bunga Iris
(Sumber: Hasil Penelitian (2021))

Teknik *Unsupervised Learning* menggunakan jenis Reduksi Dimensi (*Demensionality Reduction*)

Pada teknik *unsupervised learning* jenis reduksi dimensi pada pembuatan *machine learning*-nya hanya membutuhkan *library* sklearn. Dalam kasus ini sklearn berguna untuk memanggil dataset iris dan algoritma yang dipakai yaitu *decomposition* (PCA). Pada tahap consume model kasus ini bertujuan untuk memanipulasi data yang aslinya ada empat dimensi menjadi dua dimensi dengan menggunakan model transform. Berikut adalah *source code* yang merupakan tahap implementasi pembuatan program, dapat dilihat pada tabel 3.3a:

Tabel 3.3a. *Source Code* Tahap Implementasi *Machine Learning* menggunakan *Demensionality Reduction*

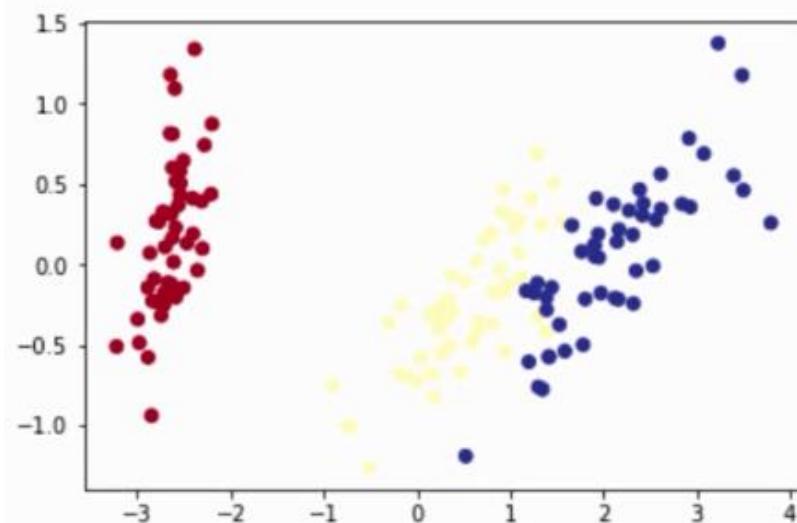
Keterangan	Source Code
Me-import library	<pre>from sklearn.datasets import load_iris from sklearn.decomposition import PCA</pre>
Menyiapkan dataset	<pre>iris = load_iris() X, y = iris.data, iris.target</pre>
Menyiapkan metode reduksi dimensi	<pre>model = PCA(n_components=2)</pre>
Me-train model	<pre>model.fit(X)</pre>
Me-consume model	<pre>X_red = model.transform(X)</pre>

Dari langkah-langkah di atas, apabila menginginkan *plotting* (visualisasi) dari hasil perograman di atas maka diperlukan langkah pada tabel 3.3b. berikut:

Tabel 3.3b. *Source Code* Tahap Visualisi Hasil menggunakan *Demensionality Reduction*

Keterangan	Source Code
Memvisualisasikan hasil	<pre>plt.scatter(X_red[:, 0], X_red[:, 1], c=y, cmap='RdY1Bu')</pre>

Berdasarkan pengimplementasian *sorce code* pada tabel 3.3a dan 3.3b di atas, maka akan menghasilkan *output* visualisasi ketika di-*Run Cells*, yaitu akan menghasilkan tampilan yang berbeda dengan tampilan sebelumnya yaitu menampilkan hasil yang merupakan hasil reduksi dari keempat fitur tersebut yang mana angka-angka tersebut diperoleh dari proses *train* model, sebagaimana bisa dilihat pada gambar 3.4 berikut:



Gambar 3. 4 Hasil Visualisasi Reduksi Dimensi Bunga Iris
(Sumber: Hasil Penelitian (2021))

Teknik *Unsupervised Learning* menggunakan jenis pengelompokan (*clustering*)

Pada teknik *unsupervised learning* jenis reduksi dimensi pada pembuatan *machine learning*-nya hanya membutuhkan library sklearn saja. Sklearn di kasus ini mempunyai fungsi untuk memanggil dataset iris dan algoritma *clustering* yaitu *K-Means*. Berikut adalah *source code* yang merupakan tahap implementasi pembuatan program dan disini jenis *clustering* akan memanfaatkan code dari jenis reduksi dimensi sebagai bandingan, dapat dilihat pada tabel 3.4a:

Tabel 3.4a. *Source Code* Tahap Implementasi *Machine Learning* menggunakan *K-Means*

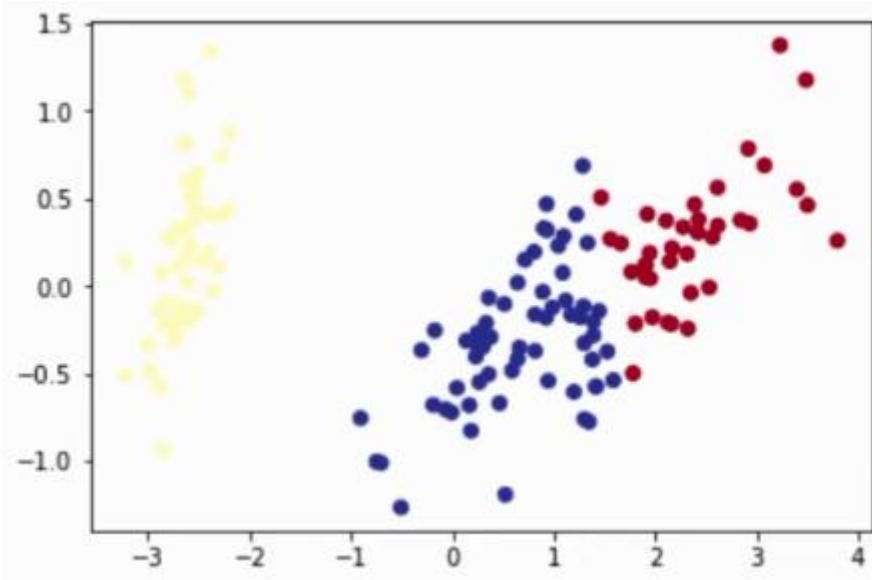
Keterangan	Source Code
Me-import library	<pre>from sklearn.datasets import load_iris from sklearn.cluster import KMeans</pre>
Menyiapkan dataset	<pre>iris = load_iris() X, y = iris.data, iris.target</pre>
Menyiapkan metode <i>clustering</i>	<pre>model = KMeans(n_clusters=3)</pre>
Men-train model	<pre>model.fit(X)</pre>
Men-consume model	<pre>pred = model.predict(X)</pre>

Dari langkah-langkah di atas, apabila menginginkan penampilan plot (visualisasi) dari hasil perograman di atas maka diperlukan pengaplikasian langkah pada tabel 3.4b. berikut:

Tabel 3.4b. *Source Code* Tahap Visualisi Hasil menggunakan *K-Means*

Keterangan	<i>Source Code</i>
Memvisualisasikan hasil	<pre>plt.scatter(X_red[:, 0], X_red[:, 1], c=pred, cmap='RdYlBu');</pre>

Hasil *running* prrogram yang didasarkan implementasi pengkodean pada tabel 3.4a dn 3.4b di atas, akan memperoleh tampilan visusalisasi data program dari *machine learning* menggunakan algoritma *K-Means*, dapat dilihat pada gambar 3.5 berikut:



Gambar 3. 5 Hasil Visualisasi Clustering Bunga Iris
(Sumber: Hasil Penelitian (2021))

KESIMPULAN

Adapun kesimpulan yang didapat dari penelitian adalah bahwa proses implemetasi *machine learning* menggunakan *library* scikit-learn akan dapat mengenal dan memahami lebih dalam mengenai pembuatan pemrograman *machine learnig* pada teknik *supervised learning* dan *unsupervised learning*. Implementasi teknik *supervised learning* pada jenis klasifikasi menggunakan algoritma *K-Neighbors Classifier* dan pada jenis regresi menggunakan algoritma *Linier Regression*. Sedangkan pada teknik *unsupervised learning* implementasinya pada jenis reduksi dimensi dengan menggunakan algoritma PCA dan pada jenis klasterisasi/pengelompokan dengan menggunakan algoritma *K-Means*. Dataset yang digunakan pada penelitian ini yaitu dataset iris yang diperoleh dari *UCI Repository* yang merupakan *public dataset* yang paling populer. Dataset iris ini memiliki sampel acak bunga iris yaitu iris santosa, versicolor dan virginica. *Interface* yang digunakan yaitu *Jupiter Notebook* dengan bahasa pemrograman default python yang memiliki library bawaan yaitu library Scikit-Learn (sklearn), numPy, Pandas, matplotlib, dan berbagai dukungan algoritma yang disebutkan di atas atau software Pycharm. Dataset tersebut diolah menggunakan metode/teknik yang sesuai dengan tujuan dari pemecahan masalah dari latar belakang permasalahan kemudian diimplementasikan untuk memberikan hasil (*output*) yang diinginkan yang kemudian divisualisasikan.

DAFTAR PUSTAKA

- Ahmad, A. (2017) 'Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning', (June).
- Arifin, S. and Muktyas, I. B. (2018) 'Membangkitkan Suatu Matriks Unimodular Dengan Python', *Jurnal Derivat: Jurnal Matematika dan Pendidikan Matematika*, 5(2), pp. 1–9. doi: 10.31316/j.derivat.v5i2.361.
- Brownlee, J. (2015) *Basic Concepts in Machine Learning*, machinelearningmastery.com. Available at: <https://machinelearningmastery.com/basic-concepts-in-machine-learning/> (Accessed: 20 January 2021).
- <http://archive.ics.uci.edu/ml/datasets/Iris> (no date) *UCI Machine Learning Repository: Iris Data Set*. Available at: <http://archive.ics.uci.edu/ml/datasets/Iris> (Accessed: 20 January 2021).
- <http://repo.anaconda.com/> (no date) *Anaconda*. Available at: <http://repo.anaconda.com/> (Accessed: 20 January 2021).
- Pedregosa, F. dan Varoquaux, Ga'el, D. (2011) 'Scikit-learn: Machine Learning in Python', *Journal of Machine Learning Research*, 39(2014), pp. 2826–2830. doi: 2825-2830.
- Retnoningsih, E. dan Pramudita, R. (2020) 'Mengenal Machine Learning Dengan Teknik Supervised dan Unsupervised Learning Menggunakan Python', 7(2), pp. 156–165.
- Wahyono, T. (2018) "Fundamental of Python for Machine Learning: Dasar-Dasar Pemrograman Python untuk Machine Learning dan Kecerdasan Buatan". Yogyakarta: Gava Media.
- Wahyu, I., Tosida, E. T. dan Andria, F. (2019) "Teori dan Panduan Praktis Data Science dan Big Data". Edisi Pert, *Quality*. Edisi Pertama. Edited by Julianto. Lembaga Penelitian dan Pengabdian pada Masyarakat Universitas Pakuan.