

IMPLEMENTASI DATA MINING C4.5, Linear Regresi dan Kmeans DENGAN MENGGUNAKAN FRAMEWORK DJANGO PYTHON

Mochammad Faid¹, Ahmad Supri², Moh Sukron³

¹ Sistem Informasi, Fakultas Teknik Informatika, Universitas Nurul Jadid, Indonesia
^{2,3} Teknologi Informatika, Fakultas Teknik Informatika, Universitas Nurul Jadid, Indonesia

Info Artikel

Riwayat Artikel:

Diterima : **24-April-2023**

Direvisi : **20-Juni-2023**

Disetujui : **19-Juli-2023**

Kata Kunci:

Data Mining,
C4.5,
Kmean,
Framework Django.

ABSTRAK

Penggunaan data dalam pengambilan keputusan bisnis telah berkembang menjadi kunci sukses di banyak bisnis di era digital yang berkembang. Data mining dapat memberi kan pemahaman yang menyeluruh tentang data dan mendorong pengambilan keputusan yang lebih bijak dengan memeriksa pola-pola yang tersembunyi dan menemukan hubungan yang relevan, namun bagaimana cara menimplementasikan sebuah algoritma data mining kedalam sebuah website diperlukan 2 keahlian dasar yaitu keahlian webdeveloper dan keahlian dalam sebuah datascient, Pada penelitian ini akan membahas bagaimana mengabung dua keahlian tersebut menjadi satu yaitu Webdeveloper dan datascient dengan menggunakan 3 algoritma klasifikasi, regresi dan algoritma cluster

Keywords:

Data Mining,
C4.5,
Kmean,
Framework Django

ABSTRACT

The use of data in business decision making has grown to be the key to success in many businesses in the thriving digital era. Data mining can provide a thorough understanding of data and encourage wiser decision making by examining hidden patterns and finding relevant relationships, but how to implement a data mining algorithm into a website requires 2 basic skills, namely web developer expertise and web developer expertise. in a datascient, this study will discuss how to combine these two skills into one, namely Webdeveloper and datascient using 2 classification algorithms and cluster algorithms, the classification algorithm used is C4.5 while for clusters using KMeans with simple datasets it is expected to produce a website that can process datascient.

Penulis Korespondensi:

Mochammad Faid,
Program Studi Sistem Informasi
Universitas Nurul Jadid
Email: mfaid@unuja.ac.id

1. PENDAHULUAN

Data mining dikenal dengan nama Knowledge Discovery in Database (KDD) dalam jurnal ilmiah . Data mining merupakan alat yang memungkinkan pengguna mengakses data dalam jumlah besar secara cepat dengan mengekstraksi informasi dari kumpulan data yang sangat besar menggunakan pendekatan analisis statistik, matematis, dan kecerdasan buatan . Data mining bertujuan untuk mengidentifikasi pola dalam data dalam jumlah besar yang sebelumnya tidak diketahui untuk mengantisipasi masa depan dan membuat penilaian .

Teknik Data Mining mencakup klasifikasi, pengelompokan (clustering), regresi, asosiasi, analisis faktor, dan lain-lain. Teknik-teknik tersebut digunakan untuk mengubah data yang tidak terstruktur menjadi informasi yang dapat dipahami orang.

Data mining dapat digunakan untuk menguraikan penemuan pengetahuan didalam database . Data mining digunakan di berbagai bidang termasuk bisnis, pemasaran, keuangan, sains, dan ilmu social. Contoh aplikasi data mining yaitu identifikasi penipuan keuangan, menganalisis pasar untuk memahami perilaku pelanggan, menemukan obat baru untuk bisnis farmasi, dan memilih prospek terbaik untuk perekrutan bintang mahasiswa. Dengan adanya data mining maka akan diperoleh suatu rahasia atau permata berupa pengetahuan dari dalam kumpulan data-data yang banyak jumlahnya.

Kebanyakan Ahli data scientis kesulitan untuk memimplementasikan hasil penelitian kedalam sebuah website, dan kebalikannya seorang yang ahli dalam bidang webdeveloper kesulitan untuk mengimplementasikan webnya untuk menampilkan hasil olah dari data scientis, penelitian ini mencoba memimplementasikan sebuah algoritma data mining dalam sebuah website dengan menggunakan bahasa pemrograman python, python adalah bahasa pemrograman yang cukup fleksible Karena bias pemrograman ini bisa mengkomodasi kebutuhan terkait data scient dan webdeveloper, untuk website dalam penelitian ini menggunakan framework django dari python itu sendiri

2. METODE PENELITIAN

Teknik data yang sering digunakan yaitu klasifikasi, *clustering*, regresi, dan asosiasi. Teknik data mining akan menerapkan algoritma dan menginterpretasikan output yang dihasilkan. Teknik data mining yang digunakan dalam analisis data:

1. Klasifikasi

Klasifikasi merupakan algoritma supervised learning untuk menemukan pola atau fungsi untuk mendeskripsikan dan memisahkan suatu kelas data dengan kelas data yang lainnya. Klasifikasi dalam informasi mining yang merupakan tata cara pendidikan informasi untuk memprediksi suatu nilai dari sekelompok data. Algoritma klasifikasi hendak menciptakan sebuah sekumpulan ketentuan yang disebut dengan rule yang hendak digunakan sebagai penkalian untuk dapat memprediksi kelas dari informasi yang hendak diprediksi.

2. Clustering

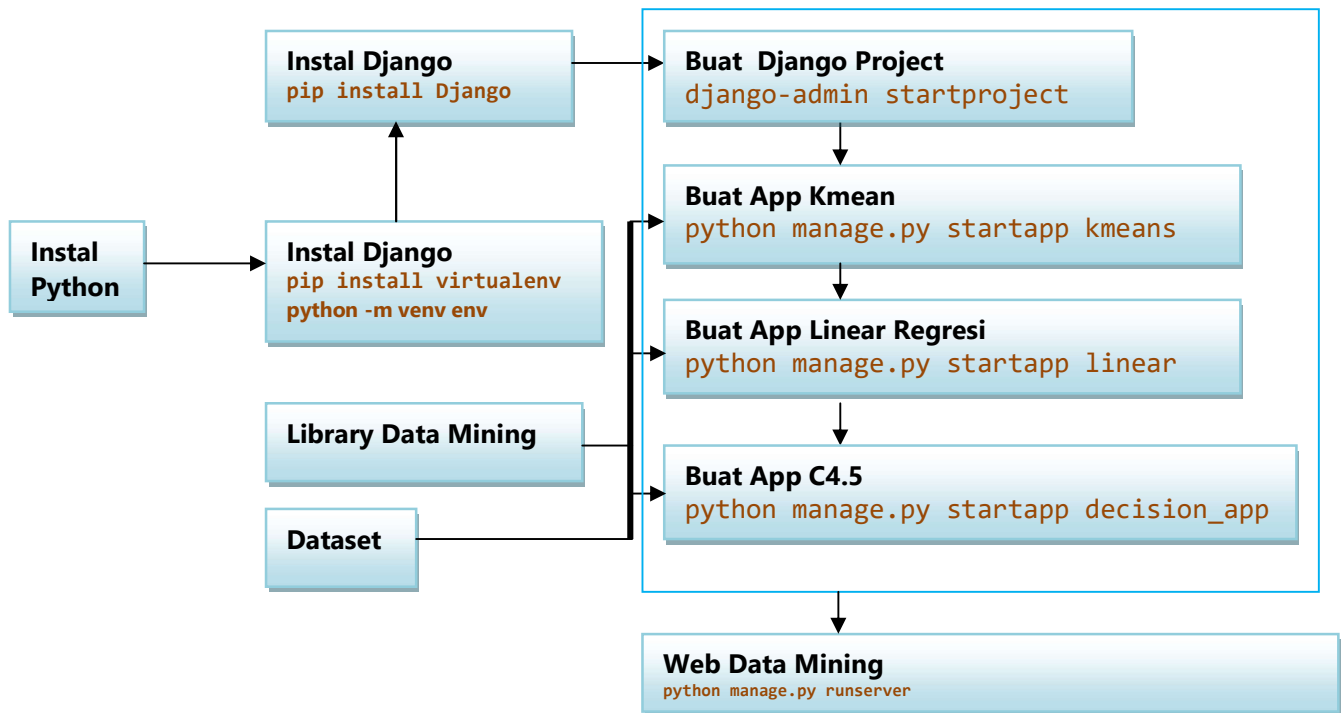
Clustering merupakan mempartisi data ke dalam teknik pengelompokan *record* pada basis data berdasarkan kriteria tertentu sehingga data yang memiliki kriteria yang sama akan dikelompokkan kedalam satu *cluster* yang sama.

3. Regresi

Analisis regresi merupakan suatu metode statistik yang mengamati hubungan antara variabel terikat Y dan serangkaian variabel bebas $X_1, \dots, X_p$. Tujuan dari metode ini adalah untuk memprediksi nilai Y untuk nilai X yang diberikan.

4. Asosiasi

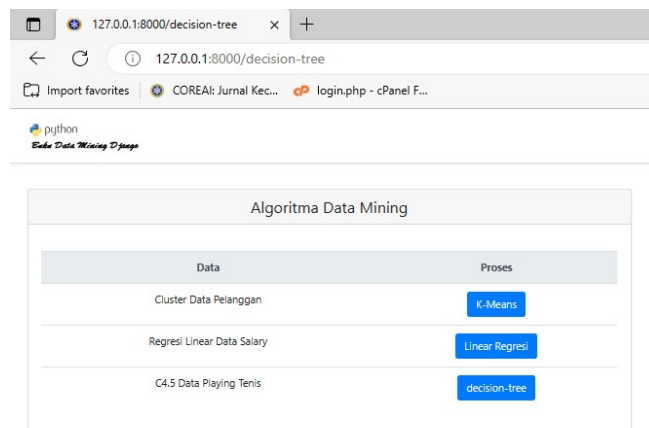
Teknik asosiasi dalam data mining merupakan teknik data mining yang dapat digunakan untuk mendapatkan pola barang yang sering dibeli secara bersamaan



Gambar 1. Metode Pembuatan Web Data Mining

3. HASIL DAN ANALISIS

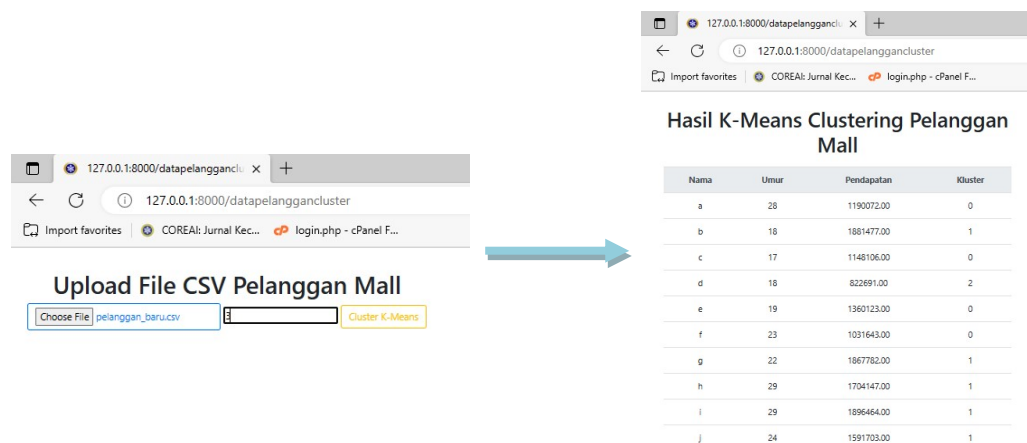
Aplikasi website data mining ini memiliki 3 algoritma data mining yang mewakili algoritma Klasifikasi, Clusteting, dan Regresi, adapun interface dari aplikasi data mining berbasis website seperti pada gambar 2 dibawah ini :



Gambar 2. Interface Aplikasi Data Mining Berbasis Website

3.1. K-Mens

K-means merupakan salah satu algoritma pembelajaran mesin yang digunakan untuk pengelompokan data (clustering) dan termasuk kedalam algoritma non hirarki yang berasal dari metode data clustering. Algoritma K-Means dimulai dari awal pembentukan partisi cluster kemudian secara iteratif partisi cluster kemudian diperbaiki hingga tidak terjadi perubahan yang signifikan pada partisi cluster.



Gambar 2. Tampilan hasil dan Cluster Kmean

Library data mining yang digunakan

```
import csv
from django.shortcuts import render
from sklearn.cluster import KMeans
from .models import Pelanggan
```

Sourcode Inti K-Means

```
def kmeans(request):
    if request.method == 'POST':
        csv_file = request.FILES['csv_file']
        k=request.POST.get('cluster_pelanggan','')
        kmeans = KMeans(n_clusters=int(k))
        kmeans.fit(data)
        labels = kmeans.labels_
        for i, p in enumerate(pelanggan):
            p.kluster = labels[i]
            p.save()
        return render(request, 'kmeans/kmeans.html', {'pelanggan': pelanggan})
    return render(request, 'kmeans/upload.html')
```

sourcecode di template html

```

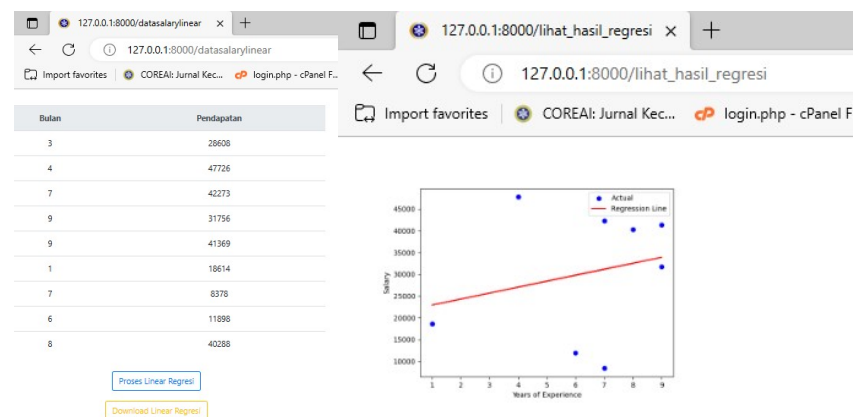
<table class="table" style="margin:20px auto;">
  <thead class="thead-light">
    <tr>
      <th scope="col">Nama</th>
      <th scope="col">Umur</th>
      <th scope="col">Pendapatan</th>
      <th scope="col">Kluster</th>
    </tr>
  </thead>

  {% for p in pelanggan %}
  <tr>
    <td>{{ p.nama }}</td>
    <td>{{ p.umur }}</td>
    <td>{{ p.pendapatan }}</td>
    <td>{{ p.kluster }}</td>
  </tr>
  {% endfor %}
</table>

```

3.2. Regresi

Metode linier ini mencari hubungan linier antara satu variabel dependen dan satu atau lebih variabel independen. Untuk menghitung regresi linear secara manual, kalian perlu menggunakan metode Ordinary Least Squares (OLS) untuk menemukan persamaan garis regresi yang terbaik



Gambar 3. Tampilan hasil dari Regresi

Library data mining yang digunakan

```

import csv
from django.shortcuts import render
from django.views import View
from matplotlib import pyplot as plt
import numpy as np

from django.http import HttpResponse, FileResponse
import os

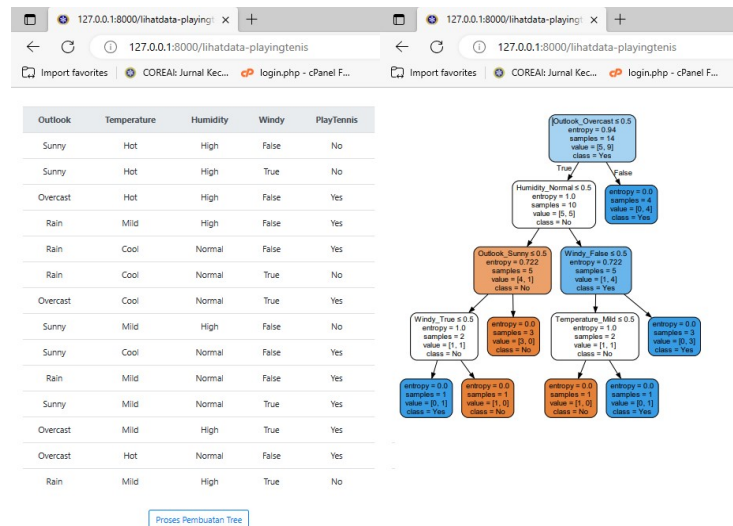
```

Sourcode Inti Linear

```
class SalaryView(View):
    template_name = 'salary/salary.html'
    csv_file = 'D:/C4.5 Django/datamining/salary/salary.csv'
    def get(self, request):
        x = []
        y = []
        with open(self.csv_file, 'r') as file:
            data = csv.reader(file)
            next(data) # Skip header
            for row in data:
                x.append(float(row[0]))
                y.append(float(row[1]))
        x = np.array(x)
        y = np.array(y)
        n = len(x)
        x_mean = np.mean(x)
        y_mean = np.mean(y)
        xy_mean = np.mean(x * y)
        x_square_mean = np.mean(x**2)
        m = (x_mean * y_mean - xy_mean) / (x_mean**2 - x_square_mean)
        c = y_mean - m * x_mean
        y_pred = m * x + c
        plt.scatter(x, y, color='blue', label='Actual')
        plt.plot(x, y_pred, color='red', label='Regression Line')
        plt.xlabel('Years of Experience')
        plt.ylabel('Salary')
        plt.legend()
        plot_path = 'D:/C4.5
Django/datamining/assets/img/hasil_linear/salary_plot.png'
        plt.savefig(plot_path)
        plt.close()
```

3.3. C4.5

Algoritma C4.5 adalah algoritma yang sangat populer untuk klasifikasi dalam machine learning dan pengolahan data[5]. Dalam algoritma C4.5 terjadi proses pengklasifikasian data dalam bentuk tingkatan yang berbeda mulai dari akar ke daun.



Gambar 4. Tampilan hasil dari C4,5

Library data mining yang digunakan

```

from django.shortcuts import render
from sklearn import tree
import pandas as pd
import matplotlib.pyplot as plt
import pydotplus
import io
import urllib, base64
import os

```

```

os.environ["PATH"] += os.pathsep + 'C:/Program Files/Graphviz/bin/'
import graphviz

```

Sourcode Inti C4.5

```

def decision_tree(request):
    # Baca dataset dari file
    data = pd.read_csv('D:/C4.5
Django/decision_tree/decision_app/playing_tennis.csv')
    data_encoded = pd.get_dummies(data, columns=['Outlook', 'Temperature',
'Humidity', 'Windy'])

    # Pisahkan fitur dan label
    X = data_encoded.drop('PlayTennis', axis=1)
    y = data_encoded['PlayTennis']

    # Buat model C4.5
    clf = tree.DecisionTreeClassifier(criterion='entropy')

    # Latih model dengan data pelatihan
    clf.fit(X, y)

    # Visualisasi pohon keputusan
    dot_data = tree.export_graphviz(clf, out_file=None,
                                    feature_names=X.columns,
                                    class_names=['No', 'Yes'],

```

```

filled=True, rounded=True,
special_characters=True)
graph = graphviz.Source(dot_data)
graph.view()
return render(request, "index.html")

```

4. KESIMPULAN

Setelah melakukan penelitian dan membahas hasil penelitian tersebut secara tuntas hingga tercipta sistem Pengolahan Data Sampling dalam bentuk program aplikasi, maka dapat diambil suatu kesimpulan sebagai berikut:

1. Penggabungan antara datascient dan web bisa dilakukan dengan bahasa pemrograman yang mendukung datascient dan webdeveloper yaitu python
2. Library yang digunakan dalam sebuah algoritma data mining berbeda beda tergantung dengan algoritma yang digunakan
3. Model yang dihasilkan dalam penelitian ini pada akhirnya digunakan untuk membuat sebuah rule atau pohon keputusan

REFERENSI

- [1] Abdurrahman, Ginanjar."Klasifikasi Penyakit Diabetes Melitus Menggunakan Adaboost Classifier." *JUSTINDO (Jurnal Sistem Dan Teknologi Informasi Indonesia)* 7, no. 1 (2022): 59–66.
- [2] Abi Islahudin, Rizal, Sidik Rahmatullah, Asep Afandi, and Sriyani Safitri."ALGORITMA C4. 5 UNTUK MEMPREDIKSI KELAYAKAN PENERIMA BANTUAN PANGAN NON TUNAI." *Jurnal Informatika* 22, no. 2 (2022): 147–59.
- [3] Adiputra, I Nyoman Mahayasa."Clustering Penyakit Dbd Pada Rumah Sakit Dharma Kerti Menggunakan Algoritma K-Means." *INSERT: Information System and Emerging Technology Journal* 2, no. 2 (2022): 99–105.
- [4] Almira, Annisa, Suendri Suendri, and Ali Ikhwan."Implementasi Data Mining Menggunakan Algoritma Fp-Growth Pada Analisis Pola Pencurian Daya Listrik." *Jurnal Informatika Universitas Pamulang* 6, no. 2 (2021): 442–48.
- [5] Apriandi, Regi, and Bitu Parga Zen."Aplikasi Terintegrasi Registrasi Torche Education Menggunakan Framework Django." *Journal of Innovation Information Technology and Application (JINITA)* 4, no. 2 (2022): 151–62.
- [6] Faid, Mochammad."Perbandingan Kinerja Tool Data Mining Weka Dan Rapidminer Dalam Algoritma Klasifikasi" 8 (2019). <https://doi.org/10.34148/teknika.v8i1.95>.
- [7] Fauzan, M Nurkamal, and Roni Habibi."Analisis Dan Perancangan Tracing Sistem Risiko Klim Pembayaran Pt Taspen (Persero) Kcu Bandung." *Jurnal Teknik Informatika* 13, no. 3 (2021): 46–55.
- [8] Handoko, Muhammad Ridho, and Neneng Neneng."Sistem Pakar Diagnosa Penyakit Selama Kehamilan Menggunakan Metode Naive Bayes Berbasis Web." *Jurnal Teknologi Dan Sistem Informasi* 2, no. 1 (2021)