Membangun Proyek Machine Learning

Walker Valentinus Simanjuntak

Institut Teknologi Del

Dicoding Indonesia

Abstrak

[The abstract should be one paragraph of between 150 and 250 words. It is not indented. Section titles, such as the word Abstract above, are not considered headings so they don’t use bold heading format. Instead, use the Section Title style. This style automatically starts your section on a new page, so you don’t have to add page breaks. Note that all of the styles for this template are available on the Home tab of the ribbon, in the Styles gallery.]

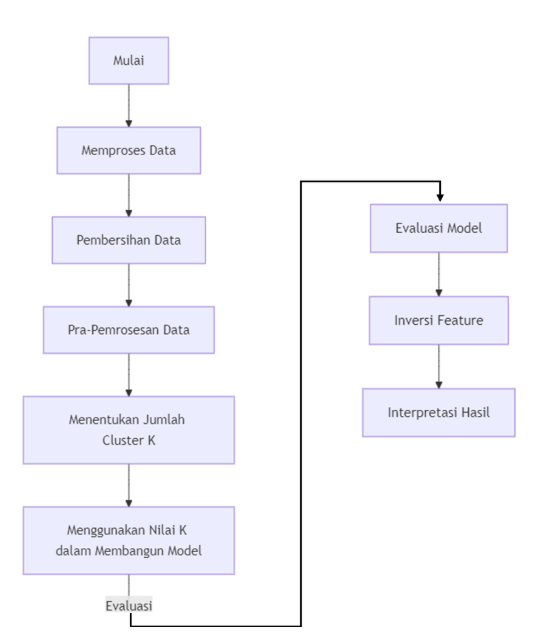
Keywords: [Click here to add keywords.]

Membangun Proyek Machine Learning

Pada proyek ini akan digunakan metode *unsupervised learning* (*clustering*) untuk menemukan pola dan secara otomatis membuat label baru untuk setiap data yang belum memiliki label. Selanjutnya, label hasil *clustering* akan digabungkan ke dalam *dataset* asli, kemudian mengubahnya menjadi data yang siap untuk klasifikasi. Terakhir, dengan dataset yang kini sudah memiliki label, akan dilatih sebuah model *supervised learning* untuk memprediksi kelas-kelas yang telah dibuat sebelumnya.

# Desain Pengerjaan

Pengerjaan proyek akan dibagi menjadi 2 metode. Metode pertama adalah menggunakan *clustering*. Tahapan pengerjaan pada metode ini dapat dilihat pada gambar x berikut.

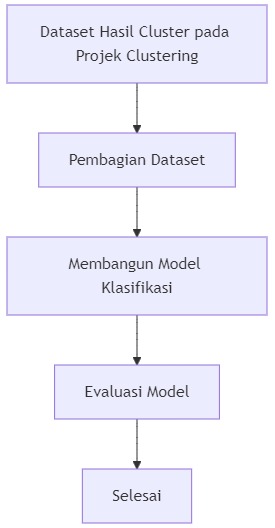


Gambar x clustering flowchart

*Flowchart* tersebut menampilkan metodologi penelitian yang diterapkan. Proses diawali dengan tahap persiapan data yang meliputi pemrosesan data mentah, pembersihan untuk menangani inkonsistensi, atau data yang hilang, serta pra-pemrosesan data seperti normalisasi atau standardisasi fitur. Langkah berikutnya adalah menentukan jumlah *cluster* optimal (K). Setelah nilai K yang paling sesuai didapatkan, nilai tersebut digunakan untuk membangun model *clustering*.

Setelah model terbentuk, alur dilanjutkan ke tahap evaluasi model untuk mengukur kualitas dan validitas *cluster* yang dihasilkan menggunakan metrik yang relevan. Untuk dapat memahami karakteristik dari setiap *cluster* secara intuitif, dilakukan proses inversi fitur (inverse transform), yang bertujuan untuk mengembalikan nilai-nilai fitur yang telah diproses ke skala aslinya. Tahapan terakhir adalah interpretasi hasil, di mana setiap *cluster* dianalisis secara mendalam untuk mendapatkan wawasan atau kesimpulan yang relevan dari pola yang ditemukan dalam data.

Metode kedua adalah menggunakan *classification*. Tahapan pengerjaan pada metode ini dapat dilihat pada gambar x berikut.



Gambar x classification flowchart

Langkah pengerjaan dimulai dari membaca dataset hasil clustering yang sudah dilakukan pada metode pertama. Setelah itu, dataset akan dibagi menjadi beberapa split yaitu train, test, dan val. Ketiga split ini akan digunakan untuk melatih (membangun) dan mengevaluasi performa model.

## Dataset

Dataset yang digunakan Adalah hasil modifikasi tim Dicoding dari dataset Kaggle “Bank Transaction Dataset for Fraud Detection” yang dilakukan untuk penilaian kriteria pemenang. Dataset terdiri dari 2537 baris dan 16 kolom.

### Heading 3

Penjelasan

#### Heading 4

Penjelasan

References

Last Name, F. M. (Year). Article Title. *Journal Title*, Pages From - To.

Last Name, F. M. (Year). *Book Title.* City Name: Publisher Name.

Footnotes

1[Add footnotes, if any, on their own page following references. For APA formatting requirements, it’s easy to just type your own footnote references and notes. To format a footnote reference, select the number and then, on the Home tab, in the Styles gallery, click Footnote Reference. The body of a footnote, such as this example, uses the Normal text style. (Note: If you delete this sample footnote, don’t forget to delete its in-text reference as well. That’s at the end of the sample Heading 2 paragraph on the first page of body content in this template.)]

Tables

Table 1

[Table Title]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column Head | Column Head | Column Head | Column Head | Column Head |
| Row Head | 123 | 123 | 123 | 123 |
| Row Head | 456 | 456 | 456 | 456 |
| Row Head | 789 | 789 | 789 | 789 |
| Row Head | 123 | 123 | 123 | 123 |
| Row Head | 456 | 456 | 456 | 456 |
| Row Head | 789 | 789 | 789 | 789 |

Note: [Place all tables for your paper in a tables section, following references (and, if applicable, footnotes). Start a new page for each table, include a table number and table title for each, as shown on this page. All explanatory text appears in a table note that follows the table, such as this one. Use the Table/Figure style, available on the Home tab, in the Styles gallery, to get the spacing between table and note. Tables in APA format can use single or 1.5 line spacing. Include a heading for every row and column, even if the content seems obvious. A default table style has been setup for this template that fits APA guidelines. To insert a table, on the Insert tab, click Table.]

Figures title:

Figure 1. [Include all figures in their own section, following references (and footnotes and tables, if applicable). Include a numbered caption for each figure. Use the Table/Figure style for easy spacing between figure and caption.]

For more information about all elements of APA formatting, please consult the APA Style Manual, 6th Edition.