**LAPORAN MACHINE LEARNING**

**GREGORIUS NOVANTINO**

1. **Bussiness Understanding :**

* **Menentukan objektif bisnis :**

Crypto menjadi investasi yang marak di kalangan anak muda. Tentunya akan sangat membantu jika ada machine learning yang mampu membantu untuk memprediksi harga dan juga menentukan pilihan apa yang harus diambil. Apakah harus menjual, hold, atau membeli koin crypto tersebut. Di prototype ini akan dilakukan ranking koin mana yang baik untuk diinvestasikan, prediksi harga koin, dan keputusan apa yang harus dilakukan.

* **Menentukan tujuan teknis :**
  + Melakukan ranking dan melihat histori data pada koin-koin yang ada di dataset
  + Memprediksi harga koin dari tanggal sekarang ke 1 minggu, 1 bulan, dan 1 tahun kedepan
  + Keputusan apa yang disarankan diambil dari prediksi harga tersebut
* **Membuat rencana proyek :** 
  + Menganalisa dataset
  + Menentukan tujuan dimanfaatkannya dataset
  + Memilih model yang cocok untuk dataset
  + Menguji coba model (trial and error)
  + Evaluasi model
  + Membuat user interface

1. **Data Understanding :**

* **Mengumpulkan data :**

Data diambil dari kaggle. Sumber :

<https://www.kaggle.com/datasets/adityamhaske/cryptocurrency-price-analysis-dataset?select=crypto_combine.csv>

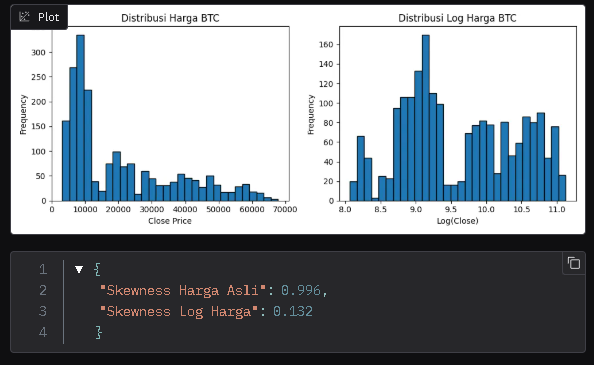
* **Menelaah & validasi data :**

Data dilakukan analisa untuk dicek apakah ada format tanggal, nama koin, atau format integer di kolom keuangan yang salah. Feature yang ditemukan :

* + Crypto : kode coin
  + Date : tanggal harian
  + Open : harga buka pasar harian
  + High : harga paling tinggi pasar harian
  + Low : harga paling rendah pasar harian
  + Close : harga tutup pasar harian
* **Contoh dataset :**



Berikut adalah hasil dari persebaran dataset :



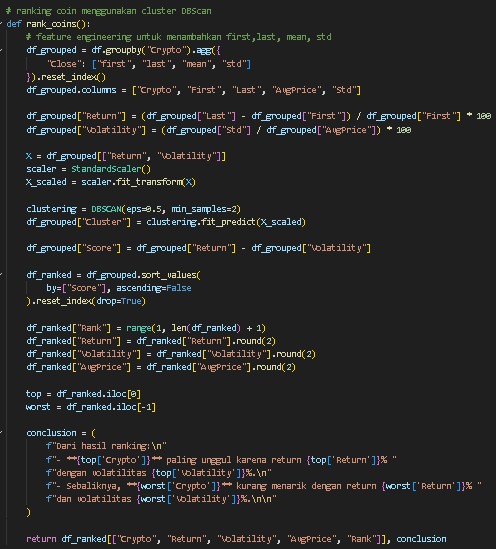
Catatan :

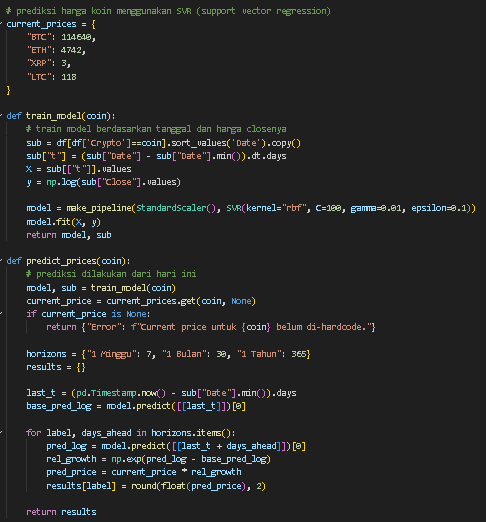
* + Harga asli : skewness dari dataset asli
  + Log harga : skewness dari dataset yang sudah diolah di training SVR

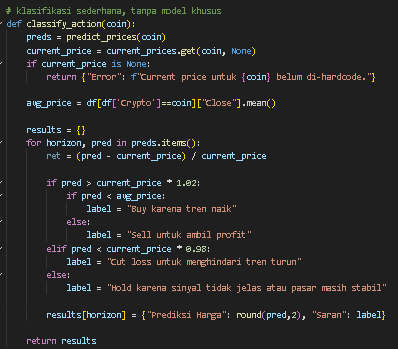
1. **Modelling :**

* **Membangun skenario pengujian & membangun model :**

Sesuai dengan yang ada di bagian bussiness understanding, ada beberapa tahapan untuk trial dan error :

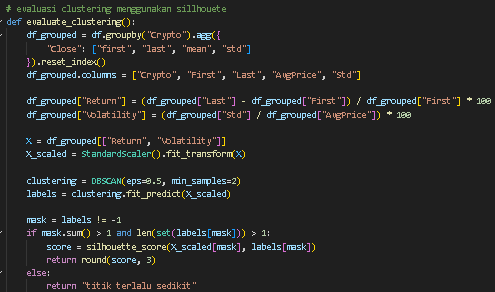
* + Membangun alur :
    - Apa yang ingin diketahui dari data?, pengambilan keputusan yang tepat untuk investasi
    - Bagaimana cara yang tepat untuk mencari keputusan itu?, dengan memprediksi harga koin
    - Apa patokan yang digunakan untuk machine learning memprediksi harga?, dari dataset yang sudah ada
  + Memilih model yang sesuai dengan alur pertanyaan :
    - Membuat informasi histori masa lalu dengan DBScan clustering untuk mengelompokkan data tiap koin dan juga ada di peringkat berapa koin tersebut berada (coin worth). Ada beberapa data yang diambil contohnya :
      * Rank : diambil dari seberapa besar return (keuntungan) dan volatility (keagresifan) suatu koin
      * Return : (Last − First) / First × 100
      * Volatility : (Std / AvgPrice) × 100
      * First : harga penutupan pertama kali (awal periode data, 2018) untuk koin itu
      * Last : harga penutupan terakhir (akhir periode data, 2023)
      * AvgPrice (mean) : rata-rata harga penutupan selama periode 2018–2023
      * Std : standar deviasi harga penutupan selama periode itu
      * Contoh kode :   
        
    - Machine learning menggunakan informasi dari dataset untuk memprediksi harga tahun 2025 atau lebih tepatnya sekarang dalam kurung waktu yang sudah ditentukan, yaitu 1 minggu, 1 bulan, dan 1 tahun kedepan. Contoh kode :



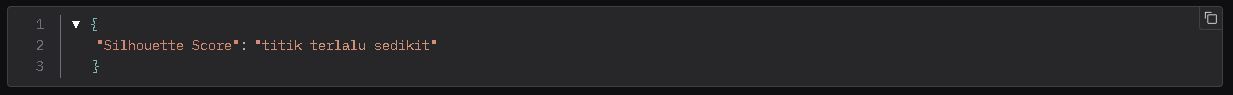
* + - Pengambilan keputusan didapat dari perhitungan return sederhana dari prediksi harga dan pelabelan berdasarkan kondisi hasil return SVR. Contoh :   
      

1. **Model Evaluation :**

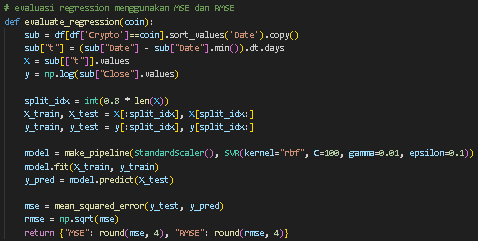
* **Evaluasi model :** 
  + Model clustering dilakukan silhouette sebagai model evaluasinya :

****

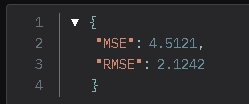
Karena jenis koin sedikit jadi silhouette tidak bisa menghitungnya :



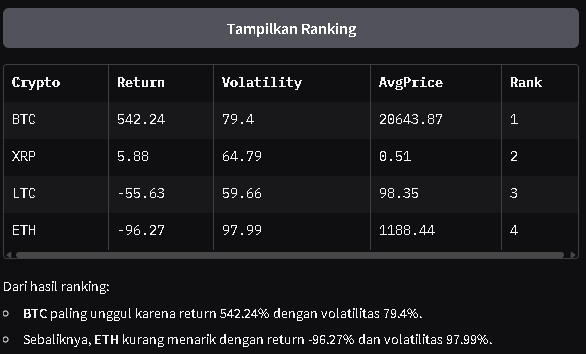
* + Model regression dilakukan RMSE dan MSE sebagai model evaluasinya :

****

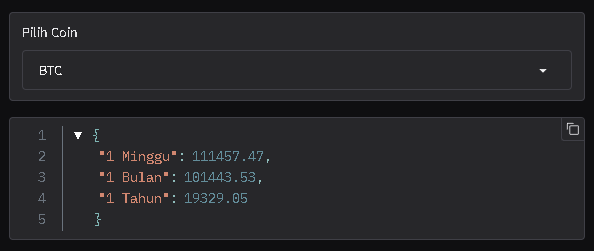
Hasil :



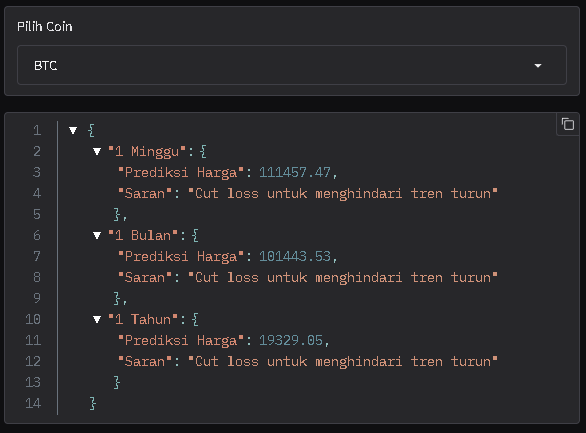
* **Evaluasi hasil model :** 
  + Setelah mencoba beberapa model clustering seperti hierarchical clustering, Kmeans clustering, dan DBScan, ternyata pemenangnya adalah DBScan karena model ini yang paling cocok dengan sifat koin crypto yang volatil dan mudah berubah harganya. Contoh hasil :



* + Setelah mencoba beberapa model regression seperti ARIMA, linear regression, polynomial regression, dan support vector regression, ternyata pemenangnya adalah support vector regression dengan hasil paling realistis. SVR juga terbukti paling stabil memprediksi dikarenakan adanya support vector dalam fungsi regresinya. Akurasi prediksi dilihat dari evaluasi regression dan data harga asli tiap koin di internet. Contoh :



* + Untuk model klasifikasi tidak ada model spesifik, hanya menggunakan rumus return biasa dan ditentukan apakah hasilnya apakah masuk di kondisi label. Contoh :



1. **Sumber Kode :**

Sumber kode dapat diakses di link github berikut :

<https://github.com/DimasAnjayMabar/proyek_ads.git>