**LAPORAN TUGAS AKHIR WORKSHOP BIG DATA**

**(Forecasting Kurs BI USD/IDR)**



**Disusun Oleh :**

* **Adi Primanto (155410139)**
* **Hari Purnomo (155410155)**
* **Nur Hidayat Pamungkas (155410111)**

**STMIK AKAKOM**

**YOGYAKARTA**

**2018**

|  |
| --- |
| **Deskripsi Masalah** |
| Pada tugas ini kami mengambil data berupa kurs history dari Bank Indonesia untuk pertukaran mata uang Dollar(USD) dan Rupiah(IDR). Data tersebut diolah dan dilakukan sebuah prediksi dalam kurun waktu satu pekan kedepan untuk mengetahui terjadinya perubahan kurs jual dari Rupiah(IDR) ke Dollar(USD). Untuk itu diperlukan analisis EDA dan menggunakan salah satu teknik dalam machine learning yaitu Forecasting dengan menggunakan metode LinearRegression. |
| **Tahap-tahap Yang Dilakukan :** |
| **1. Menentukan Data Yang Akan Diambil** |
| **2. Scraping Data** |
| import requests  from bs4 import BeautifulSoup  import csv  location = "https://pusatdata.kontan.co.id/makroekonomi/kurs\_bi"  req = requests.get(location)  soup = BeautifulSoup(req.text, 'html.parser')  scripts = soup.find\_all("script")  print soup |
| **Penjelasan:**  Pada script di atas akan mengimport library requests, BeautifulSoup, dan csv untuk menampilkan data ke dalam bentuk csv. Scraping data diperoleh dari lokasi "<https://pusatdata.kontan.co.id/makroekonomi/kurs_bi>". Pada variabel soup untuk mengambil data menggunakan BeautifulSoup. Jika soup tersebut di print maka akan menampilkan output seperti di bawah ini. |
|  |
| ranklist = soup.select('tr')  data = []  i = 0  for row in ranklist[1:]:  index = i  tanggal = row.select('td')[0].text  kurs\_jual = row.select('td')[1].text  kurs\_tengah = row.select('td')[2].text  kurs\_beli = row.select('td')[3].text  print tanggal  print kurs\_jual  print kurs\_tengah  print kurs\_beli    data.append([index, tanggal, kurs\_jual, kurs\_tengah, kurs\_beli])  print data  i=i+1  with open('databi.csv', 'w') as f:  writer = csv.writer(f)  writer.writerows(data) |
| **Penjelasan:**  Pada script di atas terdapat variabel data untuk menampung data data dan variabel I bernilai 0 dimana I tersebut nantinya untuk menampilkan data index.  Setelah itu dilakukan perulangan dan deklarasi variabel untuk pengambilan datanya.  Variabel yang pertama yaitu tanggal yang diambil dari row dengan index 0, lalu kurs jual diambil dari row dengan index 1, selanjutnya kurs tengah diambil dari row dengan index 2, selanjutnya kurs beli diambil dari row dengan index 3.  Kemudian print digunakan untuk load data dan data.append digunakan untuk menyusun data yang akan diletakkan dalam sebuah tabel.  Perintah selanjutnya membuat file dengan ekstensi .csv yang didalamnya berisi data yang telah diambil tadi. |
|  |
| import pandas as pd  kurs = pd.read\_csv('databi.csv', names=['index', 'tanggal', 'kurs\_jual', 'kurs\_tengah', 'kurs\_beli'])  kurs = kurs.sort\_values(by=['index'],ascending=False)  kurs.head(16) |
|  |
| **Penjelasan:**  Import pandas adalah sebuah librari berlisensi BSD dan open source yang menyediakan struktur data dan analisis data yang mudah digunakan dan berkinerja tinggi untuk bahasa pemrograman Python.  Selanjutnya variabel kurs dibuat untuk menampilkan data csv dalam sebuah tabel yang telah diberi nama headernya.  Kurs.head(16) merupakan perintah untuk memanggil data sebanyak 16 data. |
| **#Analisis EDA**  **#histogram kurs jual**  **plt.xlabel('kurs\_jual')**  **plt.ylabel('jumlah')**  **plt.title('Histogram Kurs Jual')**  **plt.hist(kurs['kurs\_jual'], facecolor='blue')** |
|  |
| Analisis EDA untuk Kurs Jual digunakan untuk menampilkan seluruh data kedalam bentuk histogram yang ada pada kurs jual. |
| **#histogram kurs tengah**  **plt.xlabel('kurs\_tengah')**  **plt.ylabel('jumlah')**  **plt.title('Histogram Kurs Tengah')**  **plt.hist(kurs['kurs\_tengah'], facecolor='blue')** |
|  |
| Analisis EDA untuk Kurs Tengah digunakan untuk menampilkan seluruh data kedalam bentuk histogram yang ada pada kurs tengah. |
| **#histogram kurs beli**  **plt.xlabel('kurs\_beli')**  **plt.ylabel('jumlah')**  **plt.title('Histogram Kurs Beli')**  **plt.hist(kurs['kurs\_beli'], facecolor='blue')** |
|  |
| Analisis EDA untuk Kurs Beli digunakan untuk menampilkan seluruh data kedalam bentuk histogram yang ada pada kurs beli. |
| **FORCASTING DENGAN METODE LINEAR REGRESSION** |
| **import pandas as pd**  **import numpy as np**  **import string**  **import random**  **import datetime**  **import matplotlib.pyplot as plt**  **from sklearn.linear\_model import LinearRegression**  **from sklearn import preprocessing, cross\_validation, svm** |
| **df = pd.read\_csv('databi2.csv', names=['kurs\_jual'])**  **df.head(16)** |
| Pada bagian ini dilakukan import beberapa library yang dibutuhkan untuk pengolahan data, yaitu library pandas, numpy, string, random, datetime, matplotlib.pyplot.  Selanjutnya dari sklearn.linear dilakukan import LinearRegression, prepocessing, cross\_validation dan svm.  Selanjutnya dibuat variabel df yang isinya merupakan perintah untuk memanggil data csv dan ditampilkan data kurs jual. |
| **forecast\_out = 7**  **df['Prediction'] = df[['kurs\_jual']].shift(-forecast\_out)**  **X = np.array(df.drop(['Prediction'], 1))** |
| forecast\_out merupakan jumlah angka prediksi yang akan dibuat, selanjutnya df untuk membuat tabel baru yang berisi data kurs jual. |
| **X\_forecast = X[-forecast\_out:] # set X\_forecast equal to last 30**  **X = X[:-forecast\_out] # remove last 30 from X** |
| **x\_forecast untuk mengeset data 30 terakhir, dan dilakukan pembersihan data terakhir.** |
| **y = np.array(df['Prediction'])**  **y = y[:-forecast\_out]** |
| Variabel y dibuat dengan isi merupakan perinyah untuk membuat array dan load data prediksi. |
| **X\_train, X\_test, y\_train, y\_test = cross\_validation.train\_test\_split(X, y, test\_size = 0.4)** |
| **# Training**  **clf = LinearRegression()**  **clf.fit(X\_train,y\_train)** |
| **forecast\_prediction = clf.predict(X\_forecast)**  **print(forecast\_prediction)** |
| **ts = pd.Series(data=df['kurs\_jual'])**  **ts = pd.Series(data=forecast\_prediction)**  **ts.plot()**  **plt.show** |
|  |
| Langkah terakhir yaitu menjalankan fungsi metode Linear Regression dan selanjutnya data prediksi ditampilkan dalam bentuk kurva yang terdiri dari data selama satu pekan kedepan. |