

Proses analisis sentimen dimulai dari proses pengumpulan Pengumpulan data dapat dilakukan melalui media sosial. Apabila mengumulkan data melalui media sosial. Dataset untuk analisis sentiment dapat dikumpulkan menggunakan hashtag. Selanjutnya dilakukan proses labeling bisa dilakukan berdasarkan komentar yang ada. Misalnya Topping ayamnya melimpah, dapat kita labeli dengan label Positif. Kalimat kedua : Antrinya panjang banget. Tempat agak panas penuh pembeli. Kalimat kedua dapat kita labeli dengan label negatif. Data yang sudah labeling bisa dilanjutkan ke preprocessing data. Preprocessing data meliputi pembersihan data dari karakter-karakter yang tidak relevan seperti tanda baca, stop word, serta stemming atau lemmatization untuk mengubah kata-kata menjadi bentuk dasarnya.

Data hasil preprocessing dilanjutkan proses Ekstraksi Fitur. Proses ekstraksi fitur dapat menggunakan Unigram dan TFIDF. Pengimplementasian proses ekstraksi fitur memanfaatkan class TfidfVectorizer pada library scikit-learn. Pada tahap ini, data akan diklasifikasikan menjadi tiga jenis sentimen yaitu positif, negatif, atau netral. Metode yang digunakan untuk klasifikasi sentimen dapat menggunakan machine learning. Terakhir, hasil analisis sentimen perlu dievaluasi dan disajikan secara visual agar mudah dipahami oleh pengguna.

Analisis sentimen menjadi semakin penting dalam dunia pemasaran karena memungkinkan perusahaan untuk memahami preferensi dan kebutuhan pelanggan secara lebih mendalam. Dengan mengumpulkan data dari media sosial dan platform online lainnya, perusahaan dapat menganalisis opini dan preferensi pelanggan terhadap merek mereka dan produk yang ditawarkan. Analisis sentiment dapat dilakukan salah satunya dengan menggunakan bahasa pemrograman python.

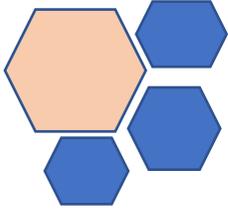


Buku Ajar

# TEORI DAN IMPLEMENTASI ANALISIS SENTIMEN MENGUNAKAN PYTHON



Yessi Yunitasari., S.Kom.M.Cs



**Buku Ajar**

**TEORI DAN IMPLEMENTASI**  
**ANALISIS SENTIMEN**  
**MENGGUNAKAN PYTHON**

**YESSI YUNITASARI, S.KOM.,M.Cs**



**Buku Ajar**  
**TEORI DAN IMPLEMENTASI ANALISIS SENTIMEN**  
**MENGGUNAKAN PYTHON**

**Penulis:**

YESSI YUNITASARI

**Perancang Sampul:**

YESSI YUNITASARI

**Penata Letak:**

Tim Kreatif UNIPMA PRESS

Cetakan Pertama, November 2023

**Diterbitkan Oleh:**

UNIPMA Press (Anggota IKAPI)

Universitas PGRI Madiun

Jl. Setiabudi No. 85 Madiun Jawa Timur 63118

Telp. (0351) 462986, Fax. (0351) 459400

E-Mail: [upress@unipma.ac.id](mailto:upress@unipma.ac.id)

Website: [kwu.unipma.ac.id](http://kwu.unipma.ac.id)

**ISBN: 978-623-8095-44-5**

Hak Cipta dilindungi oleh Undang-Undang

*All right reserved*

## KATA PENGANTAR

Segala puji bagi Allah, Tuhan Yang Maha Esa atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan buku ajar yang berjudul “**TEORI DAN IMPLEMENTASI ANALISIS SENTIMEN MENGGUNAKAN PYTHON**”. Tidak lupa juga penulis mengucapkan salawat serta salam kepada Nabi Besar Muhammad SAW, karena berkat beliau, kita mampu keluar dari kegelapan menuju jalan yang lebih terang.

Kami ucapkan juga rasa terima kasih kami kepada pihak-pihak yang mendukung lancarnya buku ajar ini mulai dari proses penulisan hingga proses cetak, yaitu orang tua kami, rekan-rekan kami, penerbit, dan masih banyak lagi yang tidak bisa kami sebutkan satu per satu.

Adapun, buku ajar telah selesai kami buat secara semaksimal dan sebaik mungkin agar menjadi manfaat bagi pembaca. Terimakasih.

Penulis

Yessi Yunitasari

# DAFTAR ISI

<b>JUDUL</b> .....	ii
<b>DAFTAR ISI</b> .....	v
<b>BAB I</b> .....	2
A. DATA MINNING .....	2
B. MANFAAT DATA MINNING.....	4
C. DATABASE SYSTEM .....	5
D. DATA WAREHOUSES.....	5
E. RINGKASAN.....	6
F. LATIHAN .....	7
<b>BAB II</b> .....	8
A. Pengumpulan Data.....	10
2.1. Labeling.....	10
2.2. Preprocessing .....	11
2.3. Ekstraksi Fitur .....	11
2.4. Training .....	11
2.5. Testing.....	11
B. RINGKASAN.....	12
C. LATIHAN .....	12
<b>BAB III</b> .....	13
A. Instalasi Python.....	13
B. Cara menulis Kode Program di Spider .....	17
C. Cara Menulis Kode Program di Goggle Colab.....	19
D. Latihan .....	25
<b>BAB IV</b> .....	26
A. Langkah Membuat Twitter Apps.....	27
B. Coding Python dengan Library Tweepy.....	31
C. Filter Berdasarkan Lokasi.....	33

D. Hasil Proses Crawling.....	36
E. Cara Crawling Data Dengan Node Js .....	37
F. Cara crawling yang diaplikasikan di Google Colabs .....	39
G. Crawling data pada Playstore.....	44
H. Latihan .....	46
<b>BAB V</b> .....	47
<b>KONSEP PREPROCESSING DATA</b> .....	47
A. Membuka Data.....	49
B. Stopword Removal.....	55
C. Penerapan stopword menggunakan corpus. ....	57
D. Case Folding .....	58
E. Tokenize.....	58
F. Memecah komentar menjadi per-kalimat. ....	59
G. Memecah komentar Menggunakan Regular Expresion. ....	59
H. Stemming .....	61
I. Import Hasil Preprocessing.....	62
J. Latihan .....	63
<b>BAB VI</b> .....	64
<b>EKSTRAKSI FITUR</b> .....	64
A. Ekstraksi Fitur Teks Menggunakan TF-IDF.....	65
B. Ekstraksi Fitur Teks Menggunakan N-Gram .....	74
C. Latihan .....	77
<b>BAB VII</b> .....	79
A. Machine Learning dengan Python .....	80
B. Melatih Machine Learning dengan Kumpulan Data.....	80
C. Algoritma dan Model Klasifikasi.....	84
<b>BAB VIII</b> .....	102
A. <i>Accuracy</i> (Akurasi) .....	104
B. Precision atau Presisi (Positive Predictive Value) .....	104

C. Recall atau Sensitivity (True Positive Rate).....	105
D. F1-Score.....	106
E. Latihan.....	107

# **BAB I**

## **PENDAHULUAN**

Pengolahan data merupakan topik yang sangat penting dalam era persaingan bisnis, dimana kita dituntut untuk menyajikan informasi dengan cepat. Informasi berkaitan dengan data untuk penentuan strategi kedepannya dan juga untuk proses bisnis sehingga penggunaan penambangan data disebut sebagai data mining.

Dengan majunya perkembangan zaman, penggunaan internet dapat menghasilkan begitu banyak data yang berjumlah sangat banyak sehingga dapat digunakan untuk mengambil keputusan. Selain itu data dapat di proses dan diolah untuk mendapatkan informasi yang bermanfaat.

### **A. DATA MINNING**

Data mining merupakan sebuah proses penyatuan data penting dan juga informasi dalam jumlah yang banyak atau big data. Terdapat beberapa metode yang dapat digunakan dalam proses pengumpulan informasi dan data, seperti penggunaan matematika, statistika, dan teknologi kecerdasan buatan (AI).

Data mining memiliki istilah lain yang biasanya dikenal sebagai *Knowledge Discovery in Database* (KDD) dan Data Analysis. Proses pengumpulan data terdapat beberapa teknik dan langkah-langkah yaitu mulai dari cleansing atau pembersihan data, integrasi data, pemilihan data dan transformasi data hingga evaluasi pola untuk mendapatkan informasi dari data tersebut.

1. Pembersihan data (untuk menghilangkan noise dan data yang tidak berubah-ubah).
2. Integrasi data (Proses penggabungan beberapa data)
3. Pemilihan data (data yang sesuai dengan tugas analisis dapat diambil dari basis data)

4. Transformasi data (data diubah dan digabungkan menjadi bentuk yang sesuai untuk penambangan dengan melakukan operasi ringkasan atau agregasi)
5. Penambangan data (sebuah proses yang penting di mana metode cerdas diterapkan untuk mengekstrak pola data)
6. Evaluasi pola (mengidentifikasi pola yang sangat menarik yang mewakili pengetahuan)
7. Presentasi pengetahuan (suatu teknik visualisasi dan penggambaran pengetahuan yang berguna dalam memaparkan pengetahuan yang diambil)

Suatu data minning menyediakan cara bagi komputer untuk mempelajari membuat keputusan data. Keputusan ini dapat memprediksi cuaca besok, suatu trend di media sosial, ataupun tanggapan masyarakat terhadap kebijakan pemerintah maupun kepuasan terhadap pelayanan publik.

Data Minning adalah bagian dari algoritma, statistik, teknik, optimisasi, dan ilmu Komputer. Data Minning juga menggunakan konsep dan pengetahuan dari bidang lain seperti linguistik, ilmu saraf, atau perencanaan kota. Penerapannya biasanya membutuhkan pengetahuan khusus untuk diintegrasikan dengan suatu algoritma tertentu.

Sebagian besar aplikasi data minning dimulai dengan membuat kumpulan data atau biasa disebut dengan Dataset. Dataset terdiri dari dari dua aspek:

1. Sampel yang merupakan objek di dunia nyata. Ini bisa berupa buku, foto, hewan, orang, daun, atau objek lainnya.

2. Fitur yang merupakan deskripsi sampel dalam dataset kami. Fitur bisa panjang, frekuensi kata tertentu, jumlah kaki, tanggal pembuatannya, dan seterusnya.

## B. MANFAAT DATA MINNING

Setiap algoritma data minning memiliki parameter, baik dalam algoritma atau disediakan oleh pengguna. Setelan ini memungkinkan sebuah algoritma untuk mempelajari cara membuat keputusan terhadap data tersebut. Sebagai contoh sederhana, komputer dapat mengkategorikan kecepatan mobil melaju tergolong “cepat” atau “lambat”. Sample data set kecepatan Laju Mobil dapat dilihat pada Tabel dibawah ini.

**Tabel 1. Tabel Kecepatan Laju Mobil**

<b>Data mobil ke -</b>	<b>Kecepatan Melaju</b>	<b>Cepat atau Lambat</b>
Mobil 1	120 km/jam	Cepat
Mobil 2	100 km/jam	Cepat
Mobil 3	45 km/jam	Lambat
Mobil 4	50 km/jam	Lambat
Mobil 5	63 km/jam	Cepat
Mobil 6	65 km/jam	Cepat
Mobil 7	80 km/jam	Cepat
Mobil 8	90 km/jam	Cepat
Mobil 9	40 km/jam	Lambat
Mobil 10	110 km/jam	Cepat

Sebuah contoh algoritma sederhana untuk data set diatas adalah jika kecepatan sebuah mobil melaju diatas 60 km/jam maka mobil tersebut

tergolong melaju dengan cepat. Sedangkan jika sebuah mobil melaju dibawah 60 km/jam maka mobil tersebut tergolong melaju dengan lambat.

### **C. DATABASE SYSTEM**

Sistem basis data atau yang biasa disebut sistem manajemen basis data (DBMS) merupakan sebuah sistem yang terdiri dari kumpulan data terkait, yang disebut basis data, dan seperangkat program perangkat lunak untuk mengelola dan mengakses data. Program dalam perangkat lunak menyediakan mekanisme untuk mendeskripsikan struktur basis data dan penyimpanan data.

Perangkat lunak berfungsi untuk pengelolaan serta menetapkan bersamaan, berbagi, serta akses data terdistribusi, memastikan konsistensi dan juga keamanan sebuah informasi dapat disimpan walaupun terdapat crash pada sistem. Database relasional merupakan kumpulan tabel yang masing - masing diberikan keunikan nama.

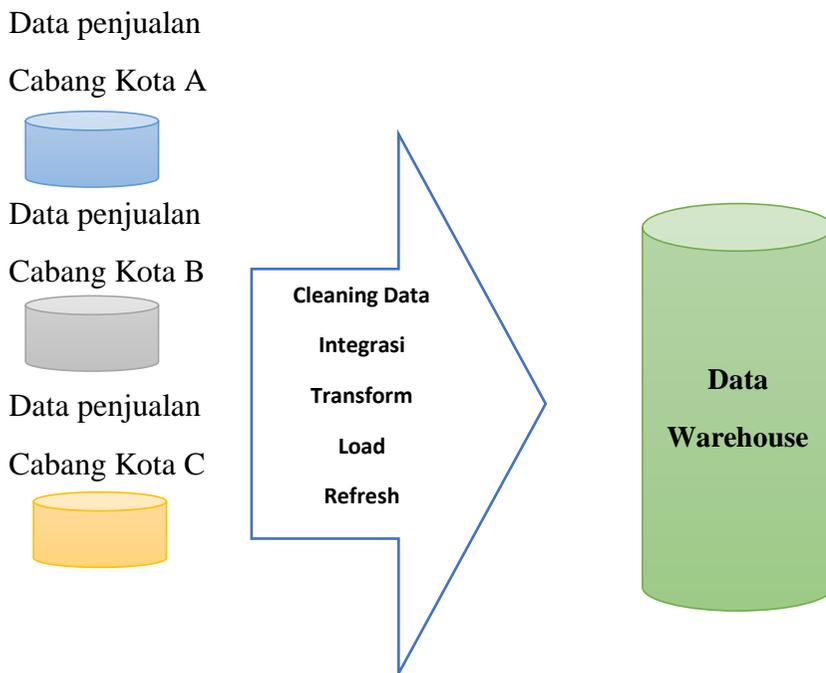
Setiap tabel terdiri dari satu set atribut (kolom atau field), biasanya disimpan dalam satu set tupel besar (record atau baris). Dalam tabel rasional unuk setiap tupel dapat mewakili objek yang diidentifikasi oleh primary key dan dijelaskan oleh satu set nilai atribut.

Model data semantik seperti model data ER atau model data entitas-hubungan, sering kali dibuat untuk database relasional. Model data entitas-hubungan mewakili database sebagai hubungan antar data dan juga satu set entitas.

### **D. DATA WAREHOUSES**

Sebuah gudang informasi yang dikumpulkan dari banyak sumber lalu disimpan dalam skema terpadu dan biasanya berada di satu situs disebut sebagai data warehouse. Data tersebut dibangun dari proses pembersihan data,

integrasi data, transformasi data, pemuatan data, dan pembaruan data secara berkala. Ilustrasi data warehouse dapat dilihat pada Gambar 1.



## E. RINGKASAN

Data Mining merupakan proses untuk menemukan pola menarik dari jumlah yang sangat besar data. Sebagai proses penemuan pengetahuan, biasanya melibatkan pembersihan data, integrasi data, pemilihan data, transformasi data, penemuan pola, evaluasi pola, dan presentasi pengetahuan. Fungsi data mining dapat digunakan untuk menentukan jenis pola atau pengetahuan meliputi karakterisasi, pengambilan pola, asosiasi, dan korelasi, klasifikasi dan regresi.

## F. LATIHAN

1. Apakah data minning itu?
2. Jelaskan langkah-langkah yang terlibat dalam data minning untuk proses penemuan pengetahuan !
3. Apakah perbedaan antara Database dan Data Warehouse?
4. Dimanakah letak kesamaan antara Database dan Data Warehouse?

## **BAB II**

### **ANALISIS SENTIMEN**

Analisis sentimen merupakan sebuah proses pengumpulan dan juga evaluasi data yang akan digunakan untuk mengidentifikasi, mengekstrak, dan memproses informasi dari berbagai sumber untuk menentukan sentimen atau pendapat yang terkait dengan topik tertentu. Adapula yang menyebutkan bahwa analisis sentiment adalah deteksi sikap-sikap (attitude) terhadap objek/orang.

Opini publik dapat menjadi faktor penting dalam keputusan bisnis, politik, dan masyarakat. Namun, seringkali sulit untuk mengukur opini publik secara akurat dan efektif. Inilah mengapa analisis sentimen menjadi begitu penting. Dengan menggunakan algoritma khusus, analisis sentimen dapat membantu kita memahami perasaan dan pandangan orang terhadap suatu topik atau produk tertentu.

Dalam konteks bisnis, analisis sentimen dapat membantu perusahaan dalam memahami umpan balik pelanggan dan menciptakan strategi yang lebih efektif untuk meningkatkan kepuasan pelanggan. Dalam bisnis, analisis sentimen sangat penting karena dapat membantu perusahaan dalam memahami preferensi dan kebutuhan pelanggan.

Analisis sentiment dapat digunakan oleh perusahaan untuk mengidentifikasi isu-isu yang mempengaruhi persepsi publik tentang merek mereka. Selain itu, analisis sentimen juga dapat membantu meningkatkan efektivitas pemasaran dengan memahami preferensi pelanggan dan menyesuaikan strategi pemasaran.

Contoh kasus penggunaan analisis sentimen dalam kehidupan nyata adalah ketika sebuah perusahaan melakukan survei untuk mengetahui bagaimana konsumen merespons produk mereka. Dengan menggunakan analisis sentimen, perusahaan dapat mengetahui apakah umpan balik dari konsumen positif atau negatif. Misalnya, jika banyak konsumen mengeluh tentang kualitas produk, perusahaan dapat memperbaiki produk mereka agar lebih memuaskan pelanggan.

Contoh berikutnya yaitu sebuah perusahaan kosmetik menggunakan analisis sentimen untuk memantau percakapan online tentang produk-produk mereka. Mereka menemukan bahwa banyak pelanggan mengeluh tentang kemasan yang sulit dibuka. Perusahaan kemudian membuat perubahan pada kemasan produk mereka, dan melihat peningkatan signifikan dalam kepuasan pelanggan. Tahapan analisis sentiment dapat dilihat pada gambar 2.1.

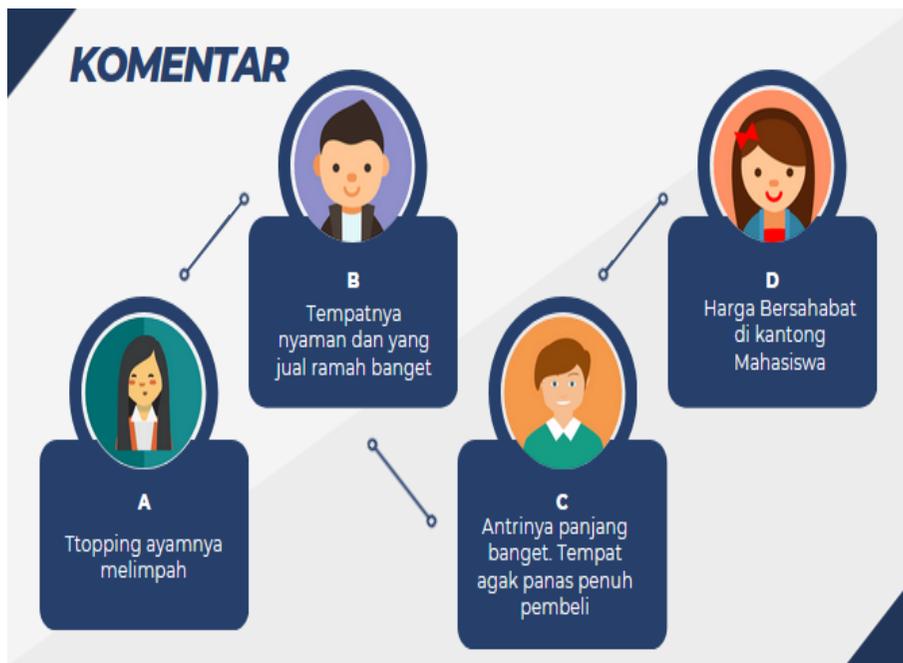


Gambar 2.1 Tahapan Analisis Sentimen

## A. Pengumpulan Data

Pengumpulan data dapat dilakukan melalui media sosial. Apabila mengumpulkan data melalui media sosial. Dataset untuk analisis sentiment dapat dikumpulkan menggunakan hashtag.

Contoh hastag yang dapat digunakan adalah #ProdukPopuler, #Bisnis, #ProdukUMKM apabila kita ingin mengumpulkan dataset terkait produk yang populer, data yang menuat kata bisnis dan data yang memuat kata produk UMKM. Apabila kita ingin mencari data terkait Tempat makan A. Maka contoh data komentar dapat dilihat pada Gambar 2.2.



Gambar 2.2 Contoh Data Komentar yang didapat berdasar hastag tertentu.

### 2.1. Labeling

Proses labeling bisa dilakukan berdasarkan komentar yang ada. Misalnya Topping ayamnya melimpah, dapat kita labeli dengan label

Positif. Kalimat kedua : Antrinya panjang banget. Tempat agak panas penuh pembeli. Kalimat kedua dapat kita labeli dengan label negatif.

## **2.2. *Preprocessing***

Preprocessing data meliputi pembersihan data dari karakter-karakter yang tidak relevan seperti tanda baca, stop word, serta stemming atau lemmatization untuk mengubah kata-kata menjadi bentuk dasarnya.

## **2.3. *Ekstraksi Fitur***

Proses Ekstraksi Fitur Menggunakan Unigram dan TFIDF. Pengimplementasian proses ekstraksi fitur memanfaatkan class `TfidfVectorizer` pada library `scikit-learn`.

## **2.4. *Training***

Pada tahap ini, data akan diklasifikasikan menjadi tiga jenis sentimen yaitu positif, negatif, atau netral. Metode yang digunakan untuk klasifikasi sentimen dapat menggunakan machine learning.

## **2.5. *Testing***

Terakhir, hasil analisis sentimen perlu dievaluasi dan disajikan secara visual agar mudah dipahami oleh pengguna.

Analisis sentimen menjadi semakin penting dalam dunia pemasaran karena memungkinkan perusahaan untuk memahami preferensi dan kebutuhan pelanggan secara lebih mendalam.

Dengan mengumpulkan data dari media sosial dan platform online lainnya, perusahaan dapat menganalisis opini dan preferensi pelanggan terhadap merek mereka dan produk yang ditawarkan.

## B. **RINGKASAN**

Menganalisis opini dan preferensi pelanggan, perusahaan dapat meningkatkan layanan pelanggan, reputasi merek, dan efektivitas pemasaran.

Perusahaan dapat mengetahui apa yang sedang dibicarakan oleh pelanggan dan bagaimana mereka merespons produk atau layanan perusahaan. Dengan adanya hal ini dapat mendukung perusahaan dalam pengambilan keputusan yang lebih baik dan akurat untuk kedepannya.

## C. **LATIHAN**

1. Apakah Analisis sentiment itu?
2. Bagaimana cara melakukan analisis sentiment untuk dunia bisnis?
3. Apa yang membedakan ketika kita akan melakukan analisis sentiment terkait dunia Pendidikan dan terkait dunia bisnis atau Industri?
4. Apa manfaat analisis sentiment untuk produsen?
5. Algoritma apa saja yang dapat digunakan untuk analisis sentiment?

## BAB III

### INSTALL PYTHON

#### A. Instalasi Python

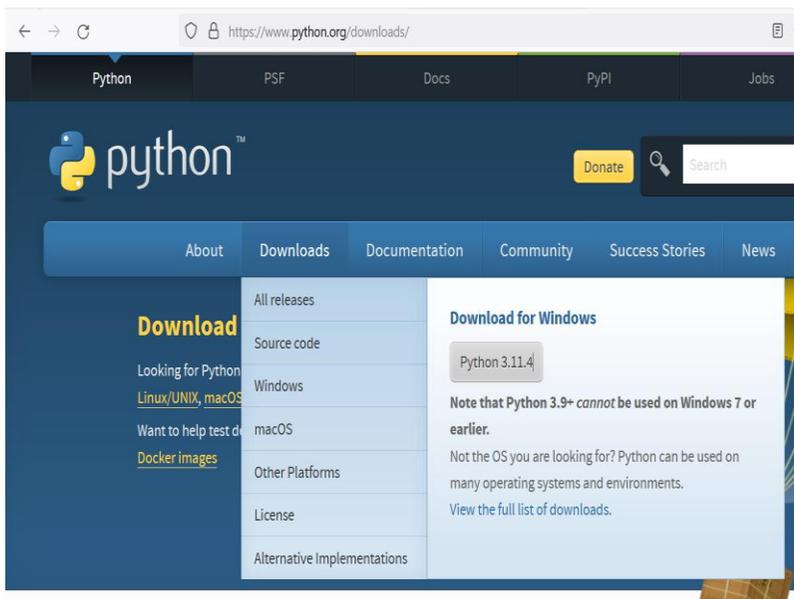
Python ialah salah satu bahasa pemrograman yang sangat populer dan juga banyak digunakan oleh para programmer untuk mengoding. Untuk para pemula juga sangat tertarik menggunakan python karena struktur sintaknya yang rapi dan juga mudah dipahami. Berikut adalah Langkah-langkah untuk install python :

1. Silhakan buka halaman ini <https://www.python.org/downloads/> .  
**Halaman yang akan ditampilkan adalah seperti ini.**



2. Selanjutnya pilihlah jenis sistem operasi sesuai yang digunakan (Windows, UNIX/Linux, Mac OS X, atau lainnya). Jika menggunakan pc atau laptop dengan OS Windows 10 maka klik Windows.

3. Klik salah satu versi Python lalu *scroll* ke bawah hingga muncul pilihan *Files* seperti gambar di atas.
4. Pilih *files* sesuai versi Windows yang digunakan yaitu 32 bit atau 64 bit. Klik file itu dan proses *download* akan mulai berjalan. Walaupun terdapat banyak pilihan/versi, kita tidak perlu bingung memilihnya. Hal yang membedakan antara satu versi dengan lainnya hanyalah fitur yang tersedia karena masing-masing versi memiliki dasar yang sama. Versi Python terbaru memiliki fitur yang paling lengkap.
5. Hal yang perlu kita perhatikan adalah ketika kita ingin mendownload python adalah versi dari OS yang kita gunakan.



Versi Python 3.9+ tidak dapat digunakan pada Windows 7 atau sebelumnya. Jadi apabila kita masih menggunakan windows 7 silahkan pilih Python versi 3.9 kebawah.

6. Setelah kita download kita dapat open file tersebut sehingga tampilannya akan muncul seperti gambar dibawah ini :



7. Kemudian kita dapat pilih Install for all users. Install for all users dipilih karena semua user computer dapat memakai python nantinya.
8. Tahap selanjutna adalah lokasi Instalasi. Tentukan lokasi dari Python akan diinstall.
9. Proses selanjutnya adalah Kostumisasi. Kostumisasi adalah proses dimana kita menentukan fitur-fitur mana saja yang akan kita install.

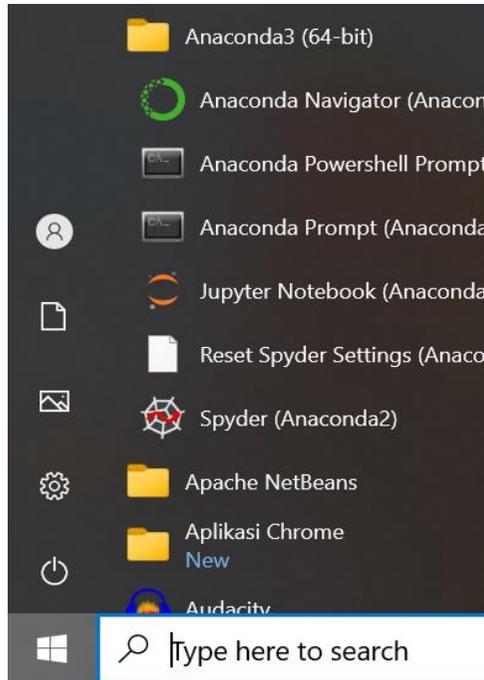


10. Aktifkan Add python.exe to path supaya CMD atau Command Prompt dapat mengenali perintah dari Python. Kemudian klik Finish untuk menyelesaikan.
11. Uji Coba Python. Setelah instalasi berhasil kita dapat uji coba python dengan membuka Python Shell. Pertama kita buka Start Menu kemudian cari **Python Shell**.
12. Cobalah ketikkan print ("Hello Semangat Belajar, Semoga Sukses"). Jika instalasi sukses maka akan muncul tampilan seperti gambar dibawah:

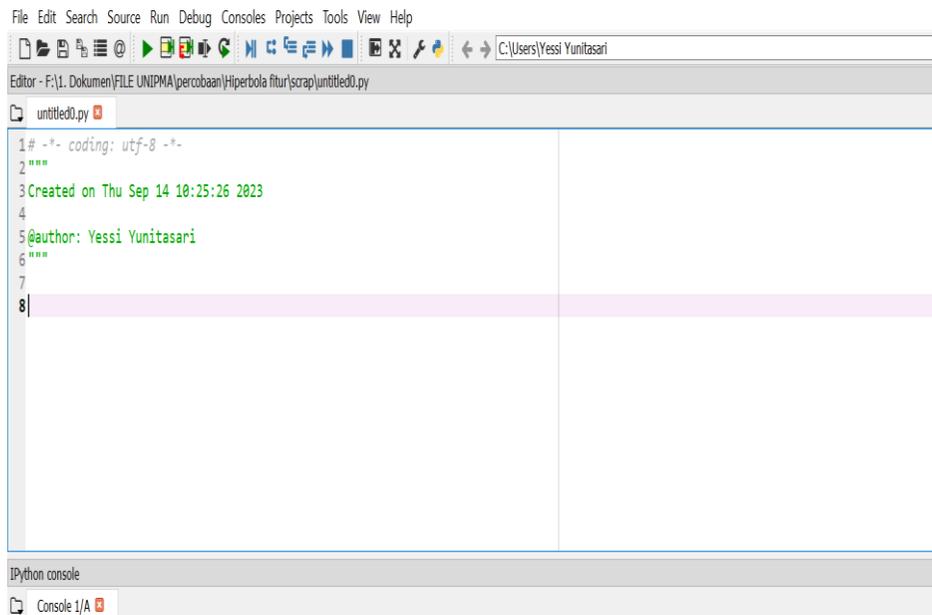
```
C:\Program Files\WindowsApps\PythonSoftwareFoundation.Python.3.10_3.10.3056.0_x64__qbz5n2kfra8p0\python3.10.11
Python 3.10.11 (tags/v3.10.11:7d4cc5a, Apr 5 2023, 00:38:17) [MSC v.1929 64 bit (AMD64)]
Type "help", "copyright", "credits" or "license()" for more information.
>>> print ("Hello Semangat Belajar, Semoga Sukses")
Hello Semangat Belajar, Semoga Sukses
>>>
```

## B. Cara menulis Kode Program di Spider

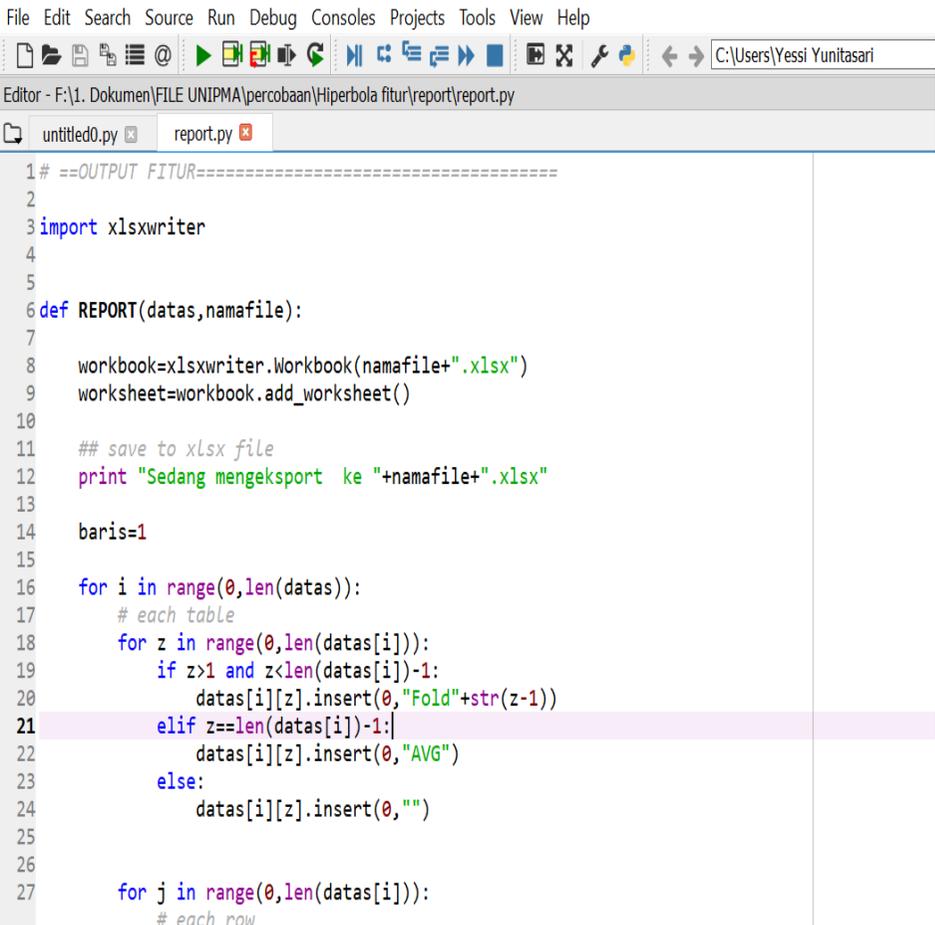
Kode program python dapat kita tuliskan pada spider. Apabila kita sudah melakukan instalasi silahkan cari spider di kompter yang sudah kita install.



Apabila sudah selesai proses open maka akan terbuka tampilan seperti di bawah ini.



Kode program siap dituliskan sesuai dengan algoritma atau sistem yang ingin dibangun. Berikut ini contoh program apabila kita ingin membuat report untuk save data dalam bentuk excel.



```
1# ==OUTPUT FITUR=====
2
3import xlswriter
4
5
6def REPORT(datas,namafile):
7
8    workbook=xlswriter.Workbook(namafile+".xlsx")
9    worksheet=workbook.add_worksheet()
10
11    ## save to xlsx file
12    print "Sedang mengekspor ke "+namafile+".xlsx"
13
14    baris=1
15
16    for i in range(0,len(datas)):
17        # each table
18        for z in range(0,len(datas[i])):
19            if z>1 and z<len(datas[i])-1:
20                datas[i][z].insert(0,"Fold"+str(z-1))
21            elif z==len(datas[i])-1:
22                datas[i][z].insert(0,"AVG")
23            else:
24                datas[i][z].insert(0,"")
25
26
27        for j in range(0,len(datas[i])):
28            # each row
```

### C. Cara Menulis Kode Program di Goggle Colab

Apabila kita ingin menuliskan program secara online kita dapat menggunakan Google Colab. Namun sebelumnya dapat kita bahas terlebih dahulu kelebihan dan kekurangan apabila kita menggunakan Google Colab.