

Three Angular Mendeteksi Kalimat Spam

Disusun Oleh

Zaky Muhammad
Yusuf



Judul :

Three Angular Mendeteksi Kalimat Spam

2023,

30 Halaman, 16 cm x 23 cm

Penulis :

Zaky Muhammad Yusuf

Penyunting :

Rolly Maulana Awangga

Layout :

Zaky Muhammad Yusuf

Sanksi Pelanggaran Pasal 113

Undang-Undang Nomor 28 Tahun 2014

tentang Hak Cipta

- (1) Setiap Orang yang dengan tanpa hak melakukan pelanggaran hak ekonomi sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf i untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 1 (satu) tahun dan/atau pidana denda paling banyak Rp100.000.000 (seratus juta rupiah).
- (2) Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf c, huruf d, huruf e, dan/atau huruf f untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 3 (tiga) tahun dan/atau pidana denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).
- (3) Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf a, huruf b, huruf c, dan/atau huruf g untuk Penggunaan Secara Komersial dipidana

Kata Pengantar

dengan pidana penjara paling lama 4 (empat) tahun dan/atau pidana denda paling banyak Rp1.000.000.000,00 (satu miliar rupiah).

- (4) Setiap Orang yang memenuhi unsur sebagaimana dimaksud pada ayat (3) yang dilakukan dalam bentuk pembajakan, dipidana dengan pidana penjara paling lama 10 (sepuluh) tahun dan/atau pidana denda paling banyak Rp4.000.000.000,00 (empat miliar rupiah).

Daftar Isi

DAFTAR ISI.....	IV
DAFTAR GAMBAR	VI
DAFTAR TABEL.....	IX
DAFTAR SIMBOL.....	X
BAB I PENDAHULUAN.....	1
1.1 LATAR BELAKANG.....	1
1.2 IDENTIFIKASI MASALAH	3
1.3 TUJUAN DAN MANFAAT.....	3
1.4 RUANG LINGKUP/BATASAN PENELITIAN.....	4
1.5 SISTEMATIKA PENULISAN	4
BAB II LANDASAN TEORI	6
2.1 SISTEM	6
2.2 SISTEM PAKAR.....	8
2.3 ANALISIS SENTIMEN	9
2.4 TEXT MINING.....	9
2.5 FAKE SPAM SMS	10
2.6 KAGGLE	ERROR! BOOKMARK NOT DEFINED.
2.7 DATASETS.....	12
2.8 METODE NAIVE BAYES.....	13

Daftar Gambar

2.9	TOKENIZER	15
2.10	CASE FOLDING.....	16
2.11	STOP WORD REMOVAL.....	16
2.12	LIBRARY SCI-KIT LEARN.....	17
2.13	WEBSITE	17
2.14	HTML.....	18
2.15	CSS.....	19
2.16	PYTHON	19
2.17	DJANGO	ERROR! BOOKMARK NOT DEFINED.
BAB III	METODOLOGI PENELITIAN	211
3.1	STUDI LITERATUR DAN PEMAHAMAN.....	21
3.2	PERSIAPAN TOOLS	21
3.3	PENGUMPULAN DATA.....	25
3.4	IMPLEMENTASI.....	26
3.5	IMPLEMENTASI PROGRAM	33
3.6	PENANGANAN ERROR.....	44
BAB IV	ANALISIS DAN HASIL PEMBAHASAN	48
4.1	ANALISIS DAN PERANCANGAN SISTEM.....	48
4.2	ANALISIS SISTEM YANG AKAN DIBANGUN	52
BAB V	PENUTUP.....	56
5.1	KESIMPULAN.....	567

DAFTAR PUSTAKA

LAMPIRAN

Daftar Gambar

Gambar 3.1 Download Python.....	222
Gambar 3.2 File Python.....	233
Gambar 3.3 Installation Type Python.....	233
Gambar 3.4 Lokasi Instalasi Python.....	244
Gambar 3.5 Instalasi Lanjutan Python	244
Gambar 3.6 Finish Instalasi Python.....	25
Gambar 3.7 Alur Implementasi.....	26
Gambar 3.8 Mencari dataset	28
Gambar 3.9 Mencari dataset	28
Gambar 3.10 Mencari dataset	29
Gambar 3.11 Mendownload dataset.....	29
Gambar 3.12 tools vsc	30
Gambar 3.13 buat file function.py.....	Error! Bookmark not defined.
Gambar 3.14 file function.py	Error! Bookmark not defined.
Gambar 3.15 Masukan dataset.csv.....	Error! Bookmark not defined.
Gambar 3.16 import library	32
Gambar 3.17 import dataset.....	32
Gambar 3.18 melihat hasilnya.....	33
Gambar 3.19 run.....	33

Gambar 3.20 hasil output.....	26
Gambar 3.21 hasil output <code>get_future</code>	34
Gambar 3.22 hasil output <code>get_future_text</code>	35
Gambar 3.23 <code>get_data</code>	36
Gambar 3.24 <code>get_train_test</code>	36
Gambar 3.25 <code>text_classifier</code>	37
Gambar 3.26 <code>show_type_of_text</code>	37
Gambar 3.27 <code>show_type_of_text</code>	38
Gambar 3.28 output	38
Gambar 3.29 <code>give_type</code>	38
Gambar 3.30 <code>show give_type</code>	39
Gambar 3.31 hasil output.....	39
Gambar 3.32 <code>show all</code>	39
Gambar 3.33 hasil output.....	40
Gambar 3.34 membuat file test	40
Gambar 3.35 file <code>test_app.py</code>	41
Gambar 3.36 import library	Error! Bookmark not defined.1
Gambar 3.37 <code>test_show_type_text</code>	Error! Bookmark not defined.2
Gambar 3.38 hasil outputnya	Error! Bookmark not defined.2
Gambar 3.39 <code>test_show_type_of_text</code>	42
Gambar 3.40 hasil outputnya	43
Gambar 3.41 <code>test_type</code>	43
Gambar 3.42 hasil outputnya	43

Daftar Gambar

Gambar 3.43 error 1.....	44
Gambar 3.44 solving error	45
Gambar 3.45 output error 2.....	45
Gambar 3.46 error 3.....	46
Gambar 3.47 output error 3.....	46
Gambar 3.48 solving error 3	46
Gambar 3.49 Hasil.....	47
Gambar 4.1 Website Bea Cukai Pengaduan.....	79
Gambar 4.2 Opsi Pengaduan	50
Gambar 4.3 Form Pertanyaan / Aduan.....	50
Gambar 4.4 Flowmap Prosedur Layanan Pengaduan.....	51
Gambar Cover (Sumber: Canva)	
Gambar Lampiran - Data Crawling (1)	



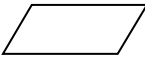
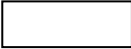
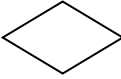
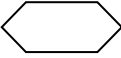



Daftar Tabel

Tabel 4.1 Deskripsi Perangkat Keras Server . **Error! Bookmark not defined.**

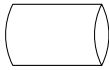

Tabel 4.2 Deskripsi Perangkat Keras Klien... **Error! Bookmark not defined.**

Tabel 4.3 Deskripsi Perangkat Lunak Server53

Daftar Simbol

No.	Simbol	Nama	Keterangan
1		Terminal / Terminator	Menunjukkan awal atau akhir dari aliran proses.
2		Arrows	Menunjukkan arus data antar simbol / proses.
3		Data	Menunjukkan data yang menjadi input / output proses.
4		Process	Menunjukkan kegiatan proses dari operasi program komputer.
5		Decision	Menunjukkan pilihan yang akan dikerjakan atau keputusan yang harus dibuat dalam proses pengolahan data.
6		Preparation	Pemberian nilai awal suatu variable.
7		Connector (On-page connector)	Digunakan untuk menghubungkan dalam satu halaman.
8		Connector (Off-page connector)	Digunakan untuk menghubungkan berbeda halaman.
9		Document	Menunjukkan dokumen sebagai yang digunakan untuk merekam data

Daftar Simbol

			terjadinya suatu transaksi.
10		<i>Database</i>	Menyimpan data berbasis <i>database</i> .
11		<i>Manual Operation</i>	Menunjukkan proses yang dikerjakan secara manual.

Bab I

Pendahuluan

1.1 Latar Belakang

Layanan pesan singkat SMS (Short Message Service) yang terdapat di ponsel genggam yang bisa kita gunakan untuk mengirim pesan singkat. SMS menggunakan tipe data pesan asinkron tempat pengiriman data dilakukan dengan penyimpanan dan mekanisme maju di mana penerima tidak perlu masuk status online, ini sangat menguntungkan bagi pengirim pesan karena tidak perlu online bersama dengan penerima pesan. Perkembangan teknologi SMS banyak disalahgunakan oleh pengguna yang tidak bertanggung jawab untuk mengirimkan pesan yang tidak berguna dan untuk dilakukan kejahatan penipuan, tentu ini sangat mengganggu kenyamanan penerima pesan. Pesan spam banyak beredar mulai dari menawarkan layanan iklan, informasi perbankan, promosi dan diskon toko dan pihak-pihak lain yang mengatas namakan provider tertentu untuk melakukan tindak kejahatan penipuan.

Penyebaran hoax masa kini dapat terbilang sangat cepat seiring dengan tingginya peningkatan penggunaan sosial media oleh masyarakat. Media sosial dapat dijangkau oleh berbagai macam usia sehingga memungkinkan semua orang untuk menjadi pengunggah atau penyebar berita di berbagai macam platform. Berita hoax dapat dikenali dengan ciri-ciri berikut yaitu; sumber yang tidak jelas, judulnya provokatif, terlihat

menggiring opini, memanipulasi fakta, berisi kebencian, dan seterusnya. Tetapi ciri-ciri tersebut tidak selalu dapat diidentifikasi secara sekilas dan harus dicari kebenarannya dari berbagai sumber serta melalui proses panjang.

Berdasarkan permasalahan di atas maka dari itu dilakukan analisis yaitu ada dua cara untuk mengatasi spam yaitu menggunakan blacklist kontak dan menggunakan metode pengklasifikasian teks. Pada penelitian kali ini yaitu membahas cara identifikasi SMS spam menggunakan metode klasifikasi bayes di mana pengklasifikasian teks berdasarkan kemungkinan kata kunci yang digunakan pada dokumen latih kemudian membandingkan dokumen latih dengan dokumen uji. yang menjadi dasar pada algoritma Bayes yaitu dengan asumsi independensi yang kuat. Klasifikasi bayes digunakan untuk memprediksi peluang masa depan yang didasari dari pengalaman sebelumnya, hal ini sangat berguna dalam mengambil keputusan yang lebih baik. Naive Bayes sangat populer dan paling banyak digunakan dalam pengklasifikasian khususnya dalam penyaringan spam sebesar 90%.

Model didalam Proses Spam Filtering bisa menggunakan berdasarkan kategori baik itu spam ataupun bukan spam (ham). untuk melakukan hal tersebut pada metode penelitian kali ini menggunakan algoritma Naive Bayes untuk mengidentifikasi secara benar dan akurat pesan yang terindikasi spam karena Naive Bayes sendiri bersifat historical (dapat mengingat kejadian sebelumnya). Hal ini sangat cocok untuk melakukan spam filtering..

1.2 Identifikasi Masalah

Berbagai masalah menjadi latar belakang permasalahan dalam hal penulisan buku *Three Angular Mendeteksi Kalimat Spam* yaitu:

1. Bagaimana cara untuk mendapatkan dataset berita yang dapat digunakan untuk training dan testing data?
2. Bagaimana mengimplementasikan metode Naïve Bayes untuk melakukan klasifikasi SMS spam?
3. Menentukan framework yang akan digunakan oleh *Sistem Identifikasi Fake News*

1.3 Tujuan dan Manfaat

Adapun tujuan dari penulisan buku *Three Angular Mendeteksi Kalimat Spam* sebagai berikut:

1. Mendapatkan dataset untuk training dan testing dengan cara download data menggunakan platform kaggle
2. Menggunakan algoritma Multinomial Naïve Bayes yang merupakan salah satu metode klasifikasi teks untuk melatih dan memprediksi data sehingga model yang didapatkan dapat diimplementasikan pada prototype *Mendeteksi Kalimat Spam*
3. Menggunakan framework Flask untuk user interface prototype *Mendeteksi Kalimat Spam*

1.4 Ruang Lingkup/Batasan Penelitian

Penggunaan Algoritma Multinomial Naïve Bayes dalam Identifikasi Mendeteksi Kalimat Spam Sms disusun dengan ruang lingkup sebagai berikut:

1. Sistem menyediakan fitur identifikasi bahwa Kalimat tersebut termasuk ke dalam sms penipuan atau tidak
2. Model memiliki akurasi 84 % sehingga layak untuk digunakan
3. Sistem menggunakan bahasa pemrograman Python dengan framework Flask, dan menggunakan metode Multinomial Naïve Bayes serta interface Jupyter Notebook
4. Analisis melewati beberapa tahapan proses yaitu download data di platform kaggle, preprocessing, labeling, data training dan testing, modeling, perancangan graphic user interface sederhana
5. Tutorial-tutorial yang dapat diikuti sebelum melakukan analisis data

1.5 Sistematika Penulisan

Laporan ini terdiri dari 5 pembahasan di antaranya :

1. Bab I Pendahuluan yang merupakan gambaran umum dari laporan, terdiri dari: latar belakang, identifikasi masalah, tujuan, ruang lingkup, dan sistematika penulisan
2. Bab II Landasan Teori yang merupakan uraian dari teori pendukung objek laporan proyek dan bersumber dari jurnal ilmiah, artikel ilmiah, buku, dll. Terdiri dari: landasan teori

3. Bab III Metodologi Penelitian yang merupakan penjelasan dari studi literatur dan pemahaman, pengumpulan data, dan implementasi.
4. Bab IV Analisis dan Hasil Pembahasan
5. Bab V Kesimpulan yang berisi kesimpulan dan saran serta maksud tujuan dari pembuatan buku
6. Bagian akhir dari sistematika penulisan yaitu daftar pustaka dan lampiran

Bab II

Landasan Teori

2.1 Sistem

Menurut Jogiyanto HM (1999:1), “Sistem adalah suatu jaringan kerja dari prosedur-prosedur yang saling berhubungan, berkumpul bersama-sama untuk melakukan suatu kegiatan atau untuk menyelesaikan suatu sasaran tertentu.”

Menurut Davis (1947), “Sistem adalah kumpulan dari elemen-elemen yang saling berinteraksi satu sama lain untuk mencapai suatu tujuan tertentu. Biasanya tujuan dari suatu sistem adalah keluaran (output) yang diinginkan oleh pembuat sistem tersebut.”

Dikutip dari McLeod (1998), “Sistem adalah tatanan (keterpaduan) yang terdiri atas sejumlah komponen fungsional (dengan satuan fungsi atau tugas khusus) yang saling berhubungan dan secara bersama-sama bertujuan untuk memenuhi suatu proses atau pekerjaan tertentu.”

Menurut Deni dan Kunkun (2015:4), “Sistem adalah kumpulan atau grup dari bagian/komponen apapun baik fisik yang saling berhubungan satu sama lain dan bekerjasama secara harmonis untuk mencapai satu tujuan.”

Romney (2005) menyatakan bahwa system merupakan suatu rangkaian dari dua atau lebih komponen-komponen yang saling berhubungan, yang berinteraksi untuk mencapai suatu tujuan.

Menurut Hall (2007) system adalah kelompok dari dua atau lebih komponen atau subsistem yang saling berhubungan yang berfungsi dengan tujuan yang sama.

The American National Standards Committee dalam La Midjan dan Susanto (2004) menyatakan bahwa system adalah system dalam pengolahan data, suatu kumpulan dari manusia, mesin, dan metode yang terorganisir untuk memenuhi seperangkat fungsi.

Sistem merupakan sekelompok unsur-unsur yang erat hubungannya satu dengan yang lain, yang berfungsi bersama-sama untuk mencapai tujuan tertentu (Sutabri, 2012).

Sistem merupakan sebuah tatanan yang terdiri atas sejumlah komponen fungsional yang saling berhubungan dan secara bersama – sama bertujuan untuk memenuhi suatu proses / pekerjaan tertentu (Kusrini, 2006).

Sistem didefinisikan sebagai suatu kumpulan atau himpunan dari unsur, komponen, atau variabel yang terorganisir, saling interaksi, saling tergantung satu sama lain, dan terpadu (Mudjahidin & Putra, 2010).

2.2 Sistem Pakar

Durkin menyatakan bahwa sistem pakar adalah suatu program komputer yang dirancang untuk memodelkan kemampuan penyelesaian masalah yang dilakukan seorang pakar.

Sistem pakar merupakan bidang studi pada Kecerdasan Buatan yang sudah ada dalam beberapa dekade (William P. Wagner, 2017).

Menurut Arhami (2005:3), sistem pakar adalah salah satu cabang yang membuat penggunaan secara luas knowledge yang khusus untuk penyelesaian tingkat manusia yang pakar.

Menurut Kusri (2008:3), sistem pakar adalah aplikasi berbasis komputer yang digunakan untuk menyelesaikan masalah sebagaimana yang dipikirkan oleh pakar. Pakar disini adalah orang yang memiliki keahlian khusus yang dapat menyelesaikan masalah yang tidak dapat diselesaikan oleh orang awam.

Menurut Syamsul (2003 : 182), sistem pakar adalah suatu sistem yang bertujuan untuk membuat keputusan yang lebih cepat daripada pakar.

Menurut McLeon (2008 : 537), sistem pakar (expert system) adalah suatu program komputer yang berusaha menampilkan pengetahuan manusia yang ahli dalam bentuk heuristic.

2.3 Analisis Sentimen

Dikutip dari Barber (2010), Sentiment Analysis adalah perpaduan dari data mining dan text mining, atau sebuah cara yang digunakan untuk mengolah berbagai opini yang diberikan oleh konsumen atau para pakar melalui berbagai media, mengenai sebuah produk, jasa ataupun sebuah instansi. Sentiment analysis merupakan sebuah metode yang digunakan untuk memahami, mengekstrak data opini, dan mengolah data tekstual secara otomatis untuk mendapatkan sebuah sentiment yang terkandung dalam sebuah opini. Pada Sentiment analysis terdiri dari 3 jenis opini, yaitu opini positif, opini negatif dan opini netral, sehingga dengan sentimen analysis perusahaan atau instansi yang terkait dapat mengetahui respon masyarakat terhadap suatu pelayanan atau produk, melalui feedback masyarakat atau pun para ahli. Sentiment mengacu pada fokus topik tertentu, pernyataan pada suatu topik mungkin akan berbeda makna dengan pernyataan yang sama pada subjek yang berbeda, oleh karena itu pada beberapa penelitian, terutama pada review produk, pekerjaan didahului dengan menentukan elemen dari sebuah produk yang sedang dibicarakan sebelum memulai proses Sentiment Analysis.

2.4 Text Mining

Dikutip dari Hearst (1997), Text mining juga dikenal sebagai data mining teks atau penemuan pengetahuan dari database tekstual, Feldman dan Dagan (1995). Sesuai dengan buku The Text Mining Handbook yang dikutip dari Feldman dan Sanger (2007), text mining dapat didefinisikan sebagai

suatu proses menggali informasi dimana seorang user berinteraksi dengan sekumpulan dokumen menggunakan tools analisis yang merupakan komponen-komponen dalam data mining. Tujuan dari text mining adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Jadi, sumber data yang digunakan dalam text mining adalah sekumpulan teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur. Adapun tugas khusus dari text mining antara lain yaitu pengkategorisasian teks dan pengelompokkan teks, Candra, dkk (2009).

Data mining merupakan proses yang memanfaatkan Teknik matematika, statistic dan kecerdasan buatan untuk mengidentifikasi informasi atau pola – pola yang valid, baru, memiliki potensi bermanfaat dan bisa dipahami dari sekumpulan data yang besar. Pola – pola tersebut bias dalam bentuk bisnis, kolerasi, trend atau model – model prediksi. Dalam proses data mining terdapat banyak langkah perulangan yang rumit yang dimana ada suatu digaan atau kesimpulan yang berbasis eksperimentasi yang dilibatkan, Albert (2019).

2.5 Fake Spam Sms

Dilansir dari Robokiller, teks spam adalah bentuk spam yang menggunakan pesan teks untuk mengirim pesan yang tidak relevan atau tidak pantas ke sejumlah besar penerima. Jika Anda telah menerima teks spam, ada kemungkinan itu ilegal..

Menurut Rubin, Chen, dan Conroy (2015) dapat disimpulkan bahwa penyebaran dan serangan hoaxes dalam skala luas dilatarbelakangi oleh teknik

yang sangat kreatif, khusus, unik dan menggunakan bermacam-macam platform yang memerlukan metode penanggulangan lebih dari hanya berdasarkan analisis teks.

Direktur Eksekutif ICT Institute (2020) ini mengatakan dengan perkembangan teknologi, ada cara mengatasi SMS spam, email, atau di WhatsApp dengan cara memblokirnya. Namun, apabila, langkah tersebut belum manjur dalam mencegah pesan tersebut, bisa mengadukannya ke regulator telekomunikasi, yang dalam hal ini Badan Regulasi Telekomunikasi Indonesia (BRTI).

H. Mooney (2018) menyatakan bahwa dari aspek sosiologi tentang fake news, ada dua kategori masalah dalam fake news yang dikelompokkan dalam post truth era dan information disorder. Ada hubungan mendasar antara sosiologi, fake news dan melek informasi. Oleh karena itu kunci permasalahan dalam fake news adalah melek informasi yang merupakan bagian dari kesadaran publik yang perlu dikembangkan dengan berangkat dari kaedah sosiologi. Dengan kaedah sosiologi kita bisa melihat masalah perseorangan sebagai masalah umum sehingga bisa diidentifikasi masalah sosial dan siklus keterkaitannya dalam skala yang lebih luas.

Nugroho (2018) menyatakan bahwa sementara itu penelitian lain menemukan bahwa ada tiga hal yang mempengaruhi tendensi seseorang menyebarkan informasi berupa hoaxes yang mencakup pondasi psikologi (psychology foundation), pondasi sosial (social foundation) dan melek media informasi (media literacy).

2.6 Kaggle

Kaggle adalah suatu situs atau platform yang mengadakan perlombaan-perlombaan di bidang Data Science. Situs ini juga menjadi salah satu sumber pembelajaran Data Science (secara praktek) yang umum.

Kaggle adalah komunitas online yang dibangun oleh Goldbloom di tahun 2010. Komunitas ini mengumpulkan para ahli ataupun minat dalam dunia data science yang ingin belajar lebih dalam terkait machine learning ataupun ilmu terkait lainnya.

Kaggle Dipergunakan untuk Data Scientist dari semua tingkatan dapat mengambil manfaat dari sumber daya dan komunitas di Kaggle. Baik jika Anda seorang pemula yang ingin mempelajari keterampilan baru dan berkontribusi pada proyek, atau seorang data scientist tingkat lanjut yang mencari kompetisi, Kaggle menjadi tempat yang tepat untuk dikunjungi.

Kaggle telah memiliki lebih dari 1000 dataset, 170.000 post di forum, dan 300 kernel. Oleh karena itu, Kaggle dianggap sebagai platform yang luar biasa hingga diakuisisi Google pada tahun 2017.).

2.7 Datasets

Dataset diartikan sebagai kumpulan data atau dokumen yang berisi satu atau lebih catatan (record). Setiap kelompok record ini tadi disebut sebagai dataset dan memiliki peran untuk menyimpan informasi seperti catatan medis, asuransi, program, dan sistem data institusi. Dataset digunakan untuk menyimpan informasi yang dibutuhkan oleh aplikasi atau sistem

operasi tertentu seperti sumber pemrograman, pustaka makro, atau variabel serta parameter sistem.(Dilansir IBM).

Menurutnya data adalah fakta kasar mengenai orang, tempat, kejadian dan sesuatu yang penting diorganisasikan. Bernard (2012 : p130)

Menurutnya data terdiri dari fakta-fakta dan angka-angka yang diolah menjadi informasi. Williams dan Sawyer (2007 : p25).

2.8 Metode Naive Bayes

Naive Bayes dikembangkan oleh Reverend Thomas Bayes pada abad ke 18. Klasifikasi dengan metode Naïve Bayes secara umum dilakukan dengan pendekatan peluang atau probabilitas. Algoritma ini memprediksi probabilitas di masa depan berdasarkan pengalaman yang sudah ada di masa lalu. Dari sudut pandang peluang, berdasarkan aturan Bayes kedalam kelas c adalah, Andre Francisco (2014):

$$P(c|d) = P(d|c)P(c)/P(d) \text{ persamaan (1)}$$

Pilihan kelas ditentukan dengan menggunakan peluang maksimal dari seluruh c dalam C , dengan fungsi:

$$\text{Argmax}_c = c \times P(d|c)(P_c)/P(d) \text{ persamaan (2)}$$

Nilai $P(d)$ konstan untuk semua kelas, maka $P(d)$ dapat diabaikan, sehingga menghasilkan fungsi:

$$F_c(d) = \text{argmax}_c E_c P(d|c)*P(c) \text{ persamaan (3)}$$

Metode yang digunakan adalah naïve bayes yaitu metode yang mempunyai perhitungan matematik dasar yang sangat kuat serta dalam efisiensi klasifikasinya juga stabil, namun kekurangannya adalah parameter model naïve bayes perlu diperkirakan dan kurang peka terhadap data yang sudah hilang. Model naïve bayes memiliki tingkat kesalahan yang sangat minimum jika dibandingkan dengan algoritma klasifikasi lainnya (Liu, Tian, Liu, Jiang, & Li, 2016). Metode naïve bayes ini merupakan salah satu metode yang populer untuk pengkategorian teks dengan frekuensi kata sebagai fitur. Hal ini dapat disimpulkan bahwa fitur-fitur yang independen dapat dibuktikan dalam algoritma klasifikasi menjadi lebih efektif (Chandrasekar & Qian, 2016).

Naïve Bayes Classifier atau disebut juga dengan Bayesian Classification merupakan metode pengklasifikasian statistik yang didasarkan pada teorema bayes yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Bayesian Classification terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database yang besar.

Naive Bayes Classifier adalah sebuah metoda klasifikasi yang berdasar pada teorema Bayes. Metode pengklasifikasian ini menggunakan metode probabilitas dan statistik yang pertama kali dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes, yaitu suatu metode untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya, sehingga metode ini dikenal sebagai Teorema Bayes. Ciri utama dari Naive Bayes Classifier ini adalah asumsi yang sangat kuat akan independensi dari masing-masing kondisi atau kejadian.

Olson dan Delen dalam bukunya, menjelaskan bahwa Naive Bayes untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar,. Algoritma ini berdasar pada mengasumsikan bahwa atribut obyek adalah sesuatu yang independen. Probabilitas yang terlibat dalam membuat perkiraan akhir dihitung sebagai jumlah frekuensi dari "master" tabel keputusan.

2.9 Tokenizer

Tokenizer merupakan proses untuk memisahkan kata atau word. Proses tokenizer dalam Datasets memiliki perbedaan dengan proses tokenizer pada teks lain. Hal ini dikarenakan adanya emoticon yang sering digunakan oleh pengguna Datasets dalam mengungkapkan perasaannya. Tahapan tokenizer dimulai dari memisah-misah bagian tweet yang dipisahkan dengan karakter spasi. Selanjutnya, bagian yang hanya memiliki satu karakter non alfabet dan angka akan dibuang. Bagian yang termasuk dalam daftar emotikon akan dikonversi. Terakhir, bagian yang memiliki karakter selain alfabet, angka, dan garis bawah akan dipecah sesuai posisi karakter tersebut.

Menurut Feldman dan Sanger (2007), Tahap tokenizing atau parsing adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Pada prinsipnya proses ini adalah memisahkan setiap kata yang menyusun suatu dokumen. Pada umumnya setiap kata teridentifikasi atau terpisahkan dengan kata yang lain oleh karakter spasi, sehingga proses

tokenizing mengandalkan karakter spasi pada dokumen untuk melakukan pemisahan kata.

2.10 Case Folding

Model kesalahan orthographic model, akan dikoreksi dengan cara mengubah semua huruf menjadi lowercase atau huruf kecil.

Kita ambil contoh adalah data tweet atau kumpulan dataset pesan spam pasti terdiri dari kalimat. Nah, untuk memudahkan proses analisis data kita harus memecah kalimat-kalimat tersebut menjadi kata atau disebut dengan token. Dengan tokenizing kita dapat membedakan mana antara pemisah kata atau bukan. Jika menggunakan bahasa pemrograman python biasanya tokenizing juga mencakup proses removing number, removing punctuation seperti simbol dan tanda baca yang tidak penting, serta removing whitespace.

2.11 Stop Word Removal

Proses ini akan menghilangkan kata-kata yang sering muncul namun tidak memiliki pengaruh apapun dalam ekstraksi sentimen suatu tweet. Kata yang termasuk kata ini adalah kata “gaul” seperti “cc”, “gw”, atau “ente”; kata hubung seperti “di”, “dengan”, atau “ke”, penunjuk waktu, dan kata tanya.

2.12 Library Sci-kit Learn

Dikutip dari Azizul (2018), Scikit-Learn merupakan suatu library untuk pengguna python pada machine learning. Scikit-learn ini merupakan free software yang dapat digunakan untuk melakukan berbagai pekerjaan dalam data science seperti regresi (regression), klasifikasi (classification), Pengelompokan (clustering), data preprocessing, dimensionality reduction dan model selection seperti pembandingan, validasi dan pemilihan parameter maupun model.

2.13 Website

Website merupakan kumpulan dari halaman-halaman yang berhubungan dengan file-file lain yang saling terkait. Dalam sebuah website terdapat satu halaman yang dikenal dengan sebutan home-page. Homepage adalah sebuah halaman yang pertama kali dilihat ketika seseorang mengunjungi sebuah website (Jhonsen, 2004 : 2).

Menurut Arief (2011:8) website adalah kumpulan dari halaman web yang sudah dipublikasikan di jaringan internet dan memiliki domain/URL (Uniform Resource Locator) yang dapat diakses semua pengguna internet dengan cara mengetikkan alamatnya.

Menurut Hidayatullah dan Kawistara (2015) “World Wide Web (WWW) adalah suatu program yang ditemukan oleh Tim Berners-Lee pada tahun 1991.”

Menurut Rohi Abdulloh (2015:1) Website atau disingkat web, dapat diartikan sekumpulan halaman yang terdiri dari beberapa laman yang berisi informasi dalam bentuk data digital baik berupa text, gambar, video, audio, dan animasi lainnya yang disediakan melalui jalur koneksi internet.

2.14 HTML

Dikutip dari Hoffer (2002) menyatakan bahwa HTML merupakan bahasa pemrograman yang digunakan untuk menampilkan dokumen pada browser dalam sebuah web. HTML bertujuan untuk mendefinisikan struktur dokumen web dan tata letak tampilan. HTML menggunakan beragam tag dan atribut. Sebuah dokumen HTML ditandai dengan tag awal dan diakhiri dengan tag.

HTML adalah Bahasa pemrograman yang fleksibel di mana kita bisa meletakkan scrip dari bahasa pemrograman lain seperti JAVA, Visual Basic, C dan lain-lain. Jika HTML tersebut tidak dapat mendukung suatu perintah pemrograman tertentu. Browser tidak akan menampilkan kotak dialog "Syntax Error" jika terdapat penulisan kode yang keliru pada scrip HTML sepanjang kode-kode yang kita tuliskan merupakan kode-kode HTML tanpa penambahan kode-kode dari luar seperti java. Oleh karena itu, jika terjadi syntax error pada skrip HTML, efek yang paling jelas adalah HTML tersebut tak akan ditampilkan pada halaman jendela browser (Rio, 2005).

2.15 CSS

Menurut Rohi Abdulloh (2015:2) CSS singkatan dari cascading style sheets, yaitu skrip yang digunakan untuk mengatur desain website. Walaupun HTML mempunyai kemampuan untuk mengatur tampilan website, namun kemampuannya sangat terbatas. Fungsi CSS adalah memberikan pengaturan yang lebih lengkap agar struktur website yang dibuat dengan HTML terlihat lebih rapi dan indah.

CSS atau singkatan dari Cascading Style Sheet adalah suatu aturan untuk mengatur tampilan dari website sehingga tampilan dalam web lebih terstruktur. CSS sendiri bukanlah bahasa pemrograman, CSS lebih seperti konfigurasi tampilan dari suatu tag pada website. CSS dapat merubah text, warna, background dan posisi dari suatu tag, Andre (2013).

2.16 Python

Pengertian Python (bahasa pemrograman) merupakan bahasa pemrograman tinggi yang bisa melakukan eksekusi sejumlah instruksi multi guna secara langsung (interpretatif) dengan metode Object Oriented Programming dan juga menggunakan semantik dinamis untuk memberikan tingkat keterbacaan syntax. Sebagai bahasa pemrograman tinggi, python dapat dipelajari dengan mudah karena telah dilengkapi dengan manajemen memori otomatis (Ilham, 2020).

2.17 Django

Django adalah sebuah framework full-stack untuk membuat aplikasi web dengan bahasa pemrograman Python. Framework akan membantu kita membuat web lebih cepat, dibandingkan menulis kode dari nol. Full-stack artinya, django meliputi sisi front-end dan juga back-end. Front-end adalah sisi depan yang akan dilihat oleh pengguna, sedangkan back-end adalah sisi belakang yang berhubungan dengan database dan logika bisnis.

Framework python ini terkenal dengan performanya yang cepat dalam mengembangkan aplikasi dan memiliki desain pragmatis yang lebih bersih. Sehingga jika Anda menggunakannya, maka proses pengembangan aplikasi menjadi lebih cepat dan tentunya menghemat kode. Pertama kali diluncurkan oleh Simon Willison dan Adrian Holovaty pada tahun 2003, Django membuat sebuah website berita. Kemudian namanya berasal dari gitaris Belgia dan Perancis yaitu Django Reinhardt. Pada bulan September 2008 rilis versi 1.0, lalu tahun 2018 sudah mencapai versi 2.0.

Django adalah sebuah web framework yang menggunakan bahasa pemrograman Python yang mendukung pembuatan website dengan konsep rapid development (Django Software Foundation, 2017).

Bab III

Metodologi Penelitian

3.1 Studi Literatur dan Pemahaman

Pada tahap ini penulis mencari dan mempelajari referensi dari berbagai *open source* seperti artikel ilmiah, contoh-contoh dataset dan program, serta jurnal nasional/internasional. Adapun topik yang akan dijelaskan ialah; analisis sentimen, *identify fake Spam Sms*, *pre-processing*, dan metode *Multinomial Naive Bayes* untuk klasifikasi.

3.2 Persiapan Tools

Environment yang digunakan untuk proses analisis data adalah jupyter notebook. Menurut website The Jupyter Notebook, jupyter notebook ialah environment komputasi interaktif, di mana kita dapat menggabungkan eksekusi kode, rich text, mathematics, plot, dan rich media.

Menurut website Jupyter Org, jupyter adalah aplikasi web asli untuk membuat dan berbagi dokumen komputasi. Ini menawarkan pengalaman yang sederhana, efisien, dan berpusat pada dokumen.

Jupyter support untuk berbagai macam bahasa pemrograman salah satunya adalah Python. Dengan minimal Python 3.3 atau versi lebih tinggi,

atau Python 2.7 untuk dapat menginstal Notebook Jupyter itu sendiri. Jupyter notebook dapat di install menggunakan Anaconda atau PIP Package manager. Berikut adalah langkah-langkah yang dapat diikuti untuk menginstal anaconda, diantaranya :

Instalasi di Windows

- Download dan Python

Pada website python.org dapat ditemukan versi terbaru sesuai dengan OS laptop/PC yang dimiliki.



Gambar 3.1 Download Python

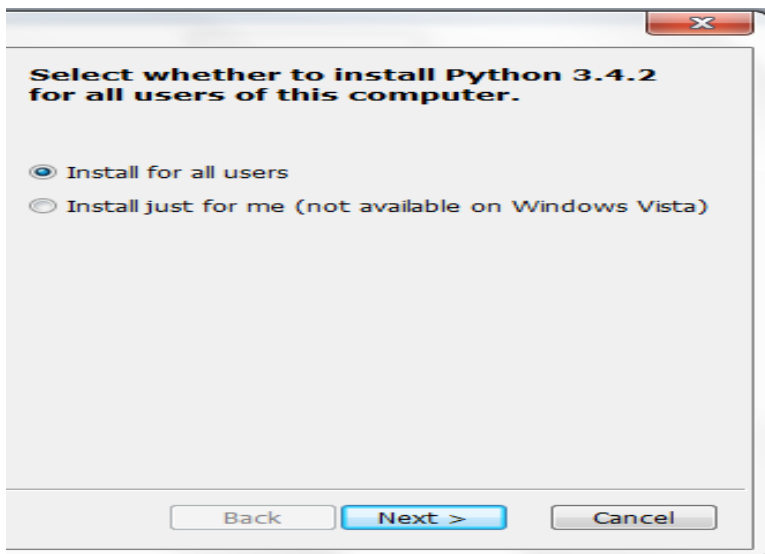
- Instalasi Python

File yang telah diunduh kemudian dibuka untuk melanjutkan ke tahap selanjutnya.



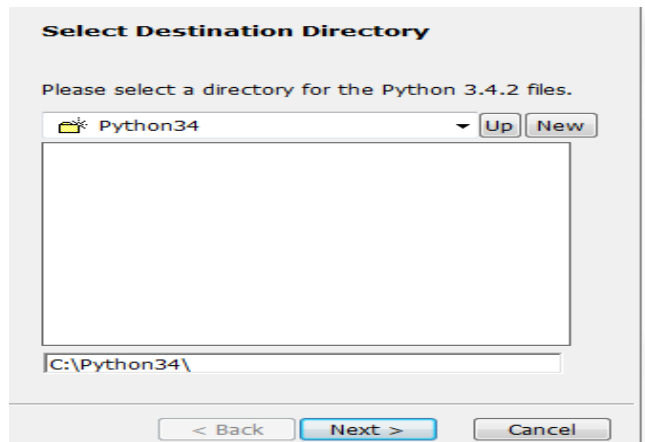
Gambar 3.2 File Python

Installation Type: **Install for all users** agar bisa dipakai untuk semua user di komputernya.



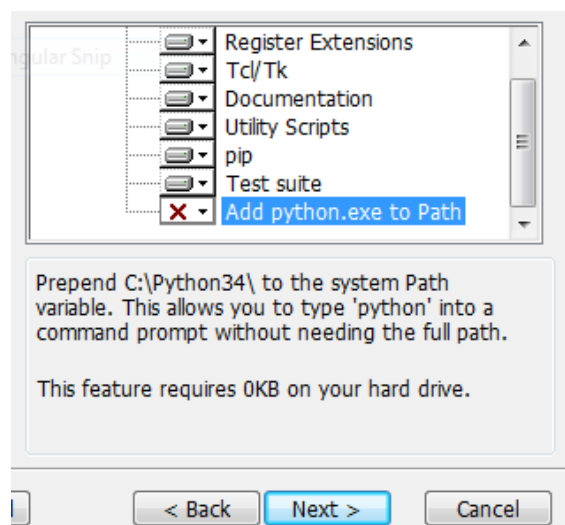
Gambar 3.3 Installation Type Python

Lokasi instalasi.



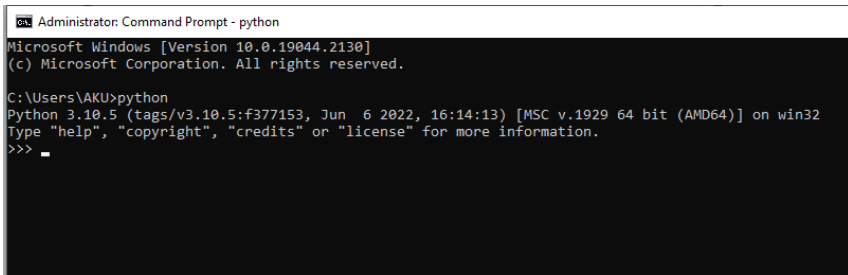
Gambar 3.4 Lokasi Instalasi Python

Opsi instalasi Kostumisasi lanjutan.



Gambar 3.5 Instalasi Lanjutan Python

Instalasi selesai.



```
Administrator: Command Prompt - python
Microsoft Windows [Version 10.0.19044.2130]
(c) Microsoft Corporation. All rights reserved.

C:\Users\AKU>python
Python 3.10.5 (tags/v3.10.5:f377153, Jun  6 2022, 16:14:13) [MSC v.1929 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>> _
```

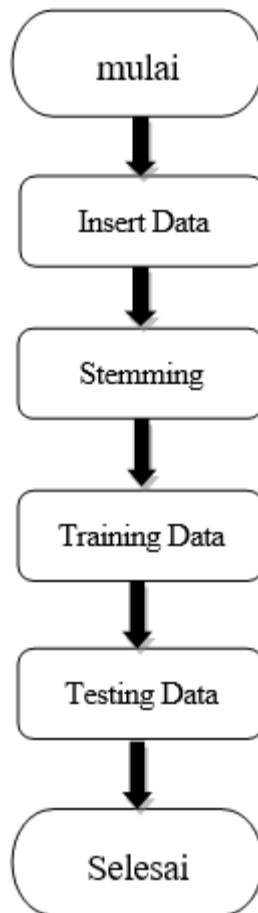
Gambar 3.6 Finish Instalasi Python

3.3 Pengumpulan Data

Dataset yang digunakan pada penelitian ialah dataset yang berisikan teks/cuitan dari berbagai media. Dataset berisikan atribut text dan tipe text nya. Didapatkan melakukan download ataupun manual pengumpulan data menggunakan Kaggle ataupun pengaduan dari forum internet. Dengan menggunakan forum dari kaggle dan kasus berita dari internet, beberapa data didapatkan kemudian disimpan sebagai file csv.

3.4 Implementasi

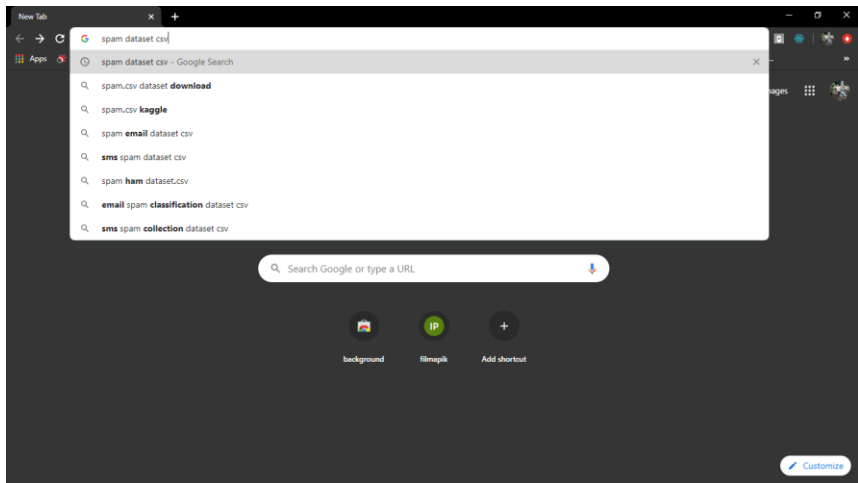
Alur implementasi pada penelitian yang dilakukan oleh penulis dapat dilihat pada diagram berikut.



Gambar 3.7 Alur Implementasi

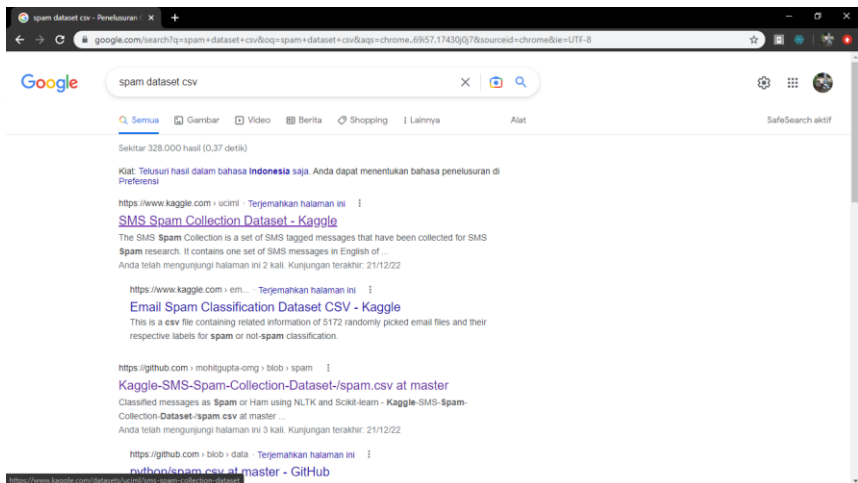
Proses awal pada proses implementasi adalah crawling data. Untuk mendapatkan dataset yang nantinya akan digunakan sebagai data train dan data test dilakukan proses crawling, yaitu proses pengambilan data menggunakan datasets yang ada di internet dan data real yang ada di pengaduan. Kemudian data disimpan dalam bentuk file csv lalu dipanggil kembali menggunakan library pandas. Selanjutnya dilakukan preprocessing data untuk menjadikan data lebih sederhana. Data dibersihkan dari character-character tidak penting, dikelompokkan berdasarkan tokennya, data yang sudah bersih kemudian disimpan sebagai file csv baru. Kemudian dilakukan labeling data berdasarkan atribut text, dengan melakukan pengecekan kebenaran terhadap isi text tersebut melalui google dan website bravo Bea Cukai khusus berita hoax. Setelah itu kemudian data labeling dengan record true atau false dipanggil kembali menggunakan pandas. Data dibersihkan kembali dengan menggunakan metode stemming, kata dikembalikan ke bentuk dasarnya. Hasil stemming kemudian digunakan sebagai dataset training. Setelah itu klasifikasi text dengan menggunakan metode Multinomial Naive Bayes.

3.5 Implementasi Program



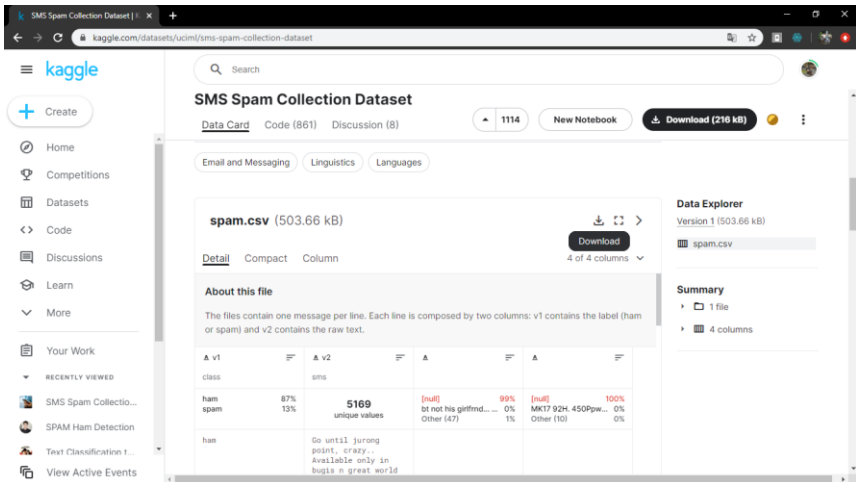
Gambar 3.8 Mencari dataset

Pertama kita mencari datasetnya di google terlebih dahulu



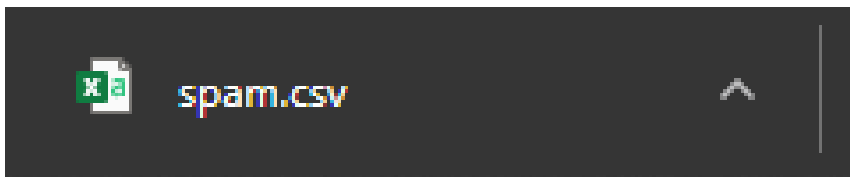
Gambar 3.9 Mencari dataset

Lalu setelah itu kita buka website nya kaggle



Gambar 3.10 Mencari dataset

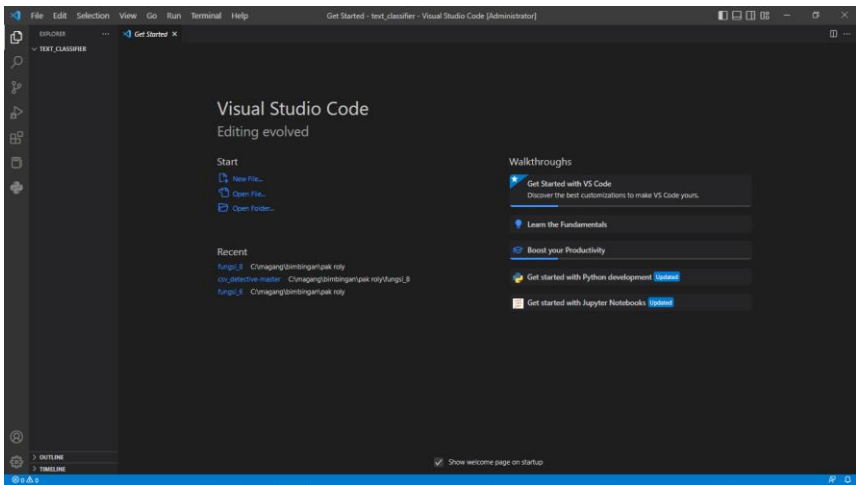
Setelah ketemu dan kita lihat sesuai sama apa yang kita butuhkan maika langsung kita download saja



Gambar 3.11 Mendownload dataset

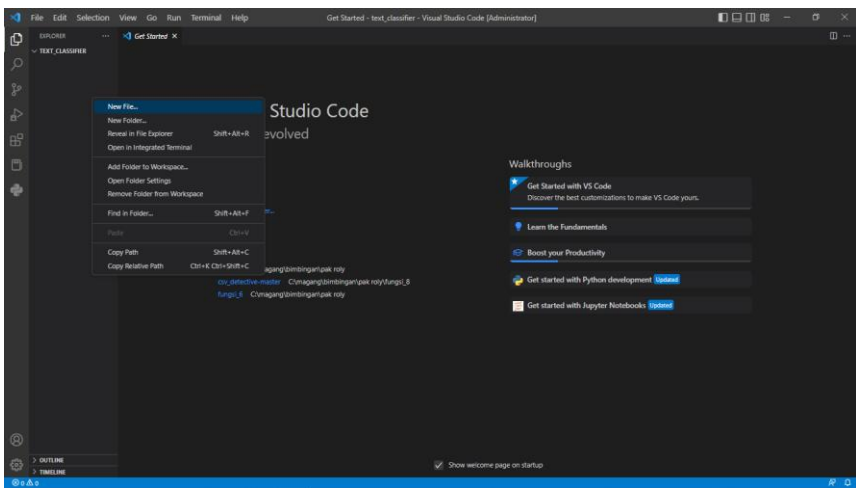
Gambar setelah di download

Bab 3 Metodologi Penelitian



Gambar 3.12 tools vsc

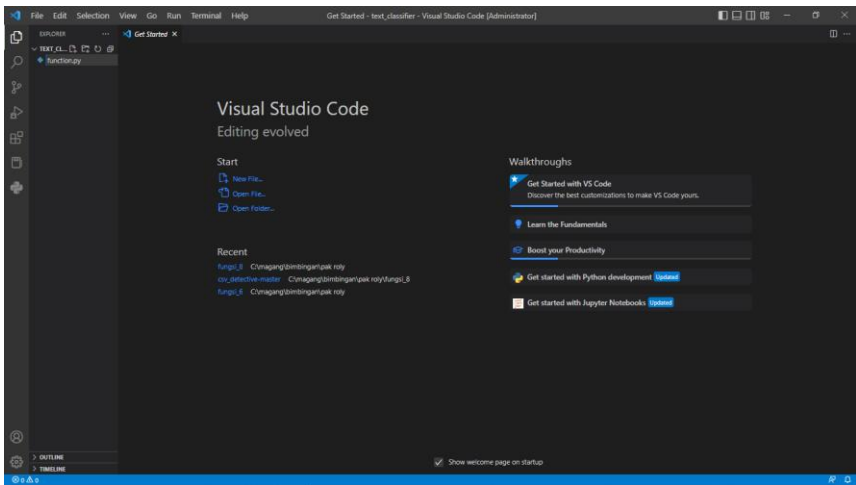
Sekarang kita bikin folder baru dan kita drag and drop di vsc



Gambar 3.13 buat file function.py

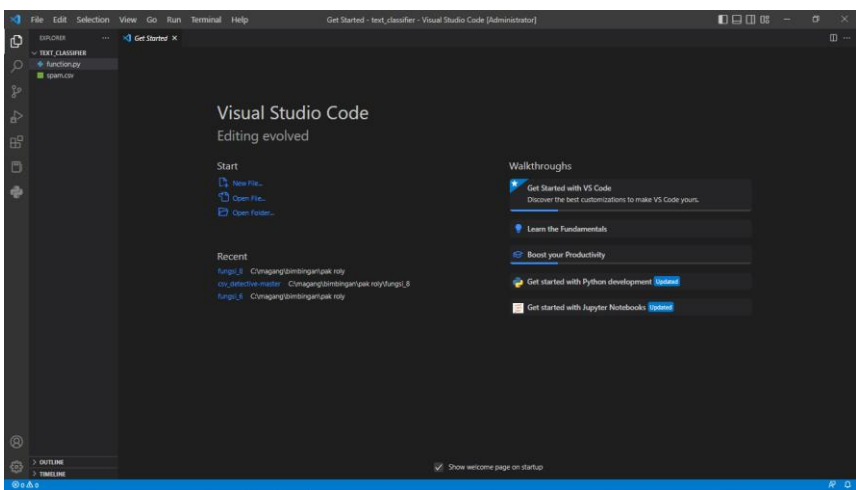
Lalu kita buat file python nya dengan nama function.py

Bab 3 Metodologi Penelitian



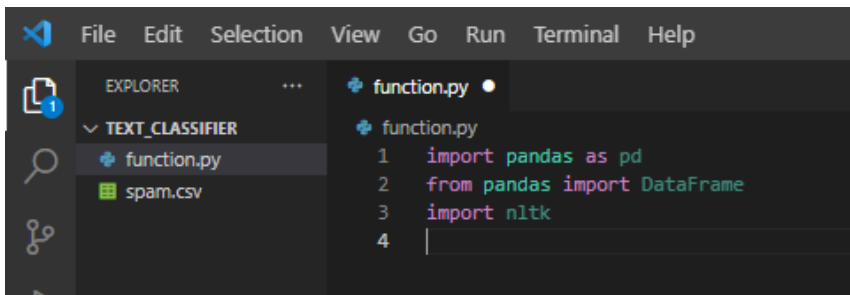
Gambar 3.14 file function.py

Menjadi seperti ini



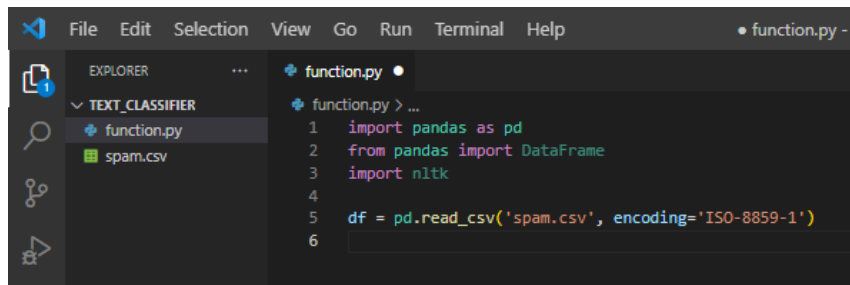
Gambar 3.15 Masukan dataset.csv

Lalu dari dataset yang kita download dari google tadi tidak lupa untuk kita masukan pada tools yang kita gunakan saat ini yaitu vsc(visual studio code)



Gambar 3.16 import library

Lalu kita masukan juga library-library yang dibutuhkan



Gambar 3.17 import dataset

Pandas merupakan library open source dengan Python yang sering digunakan untuk mengolah data yang meliputi pembersihan data, manipulasi data, hingga melakukan analisis data. Saat melakukan analisis, kami tidak dapat menggunakan data mentah.

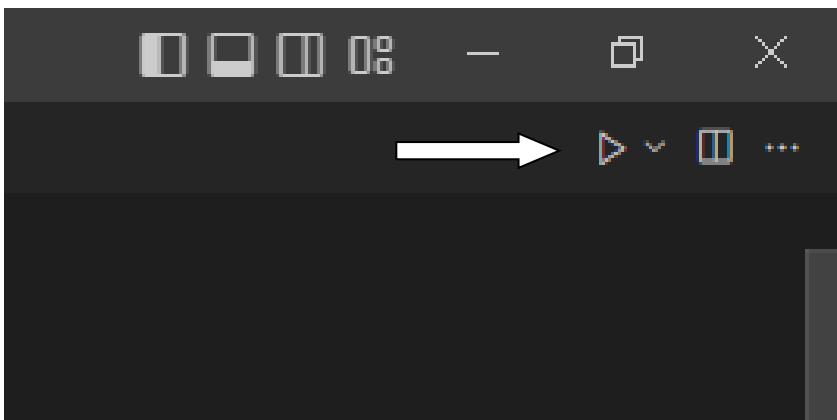
Data mentah harus diolah terlebih dahulu hingga layak untuk dianalisis. Tahap ini juga dikenal sebagai sengketa data. Perselisihan

adalah proses dimana data dikelola dan dibentuk menjadi lebih terorganisir.

```
4
5 df = pd.read_csv('spam.csv', encoding='ISO-8859-1')
6
7 if __name__ == '__main__':
8     print(df)
9
```

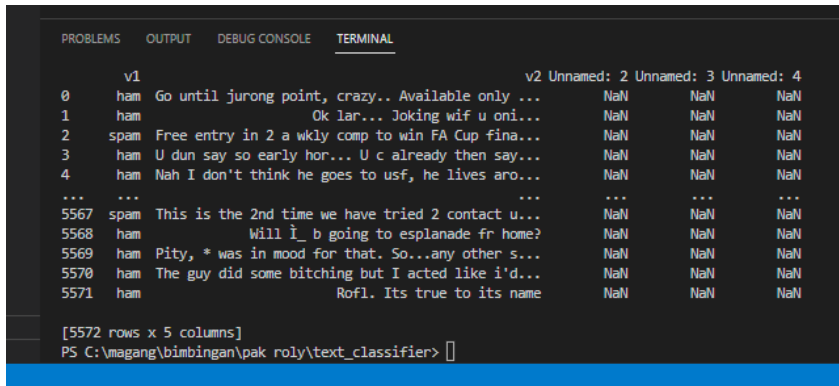
Gambar 3.18 melihat hasilnya

NLTK adalah singkatan dari Natural Language Tool Kit, yang merupakan pustaka yang digunakan untuk membantu kita bekerja dengan teks. Library ini memudahkan kita untuk mengolah teks seperti melakukan classification, tokenization, stemming, tagging, parsing, dan semantic reasoning.



Gambar 3.19 run

Disini program akan kita langsung jalan kan saja untuk melihat hasilnya



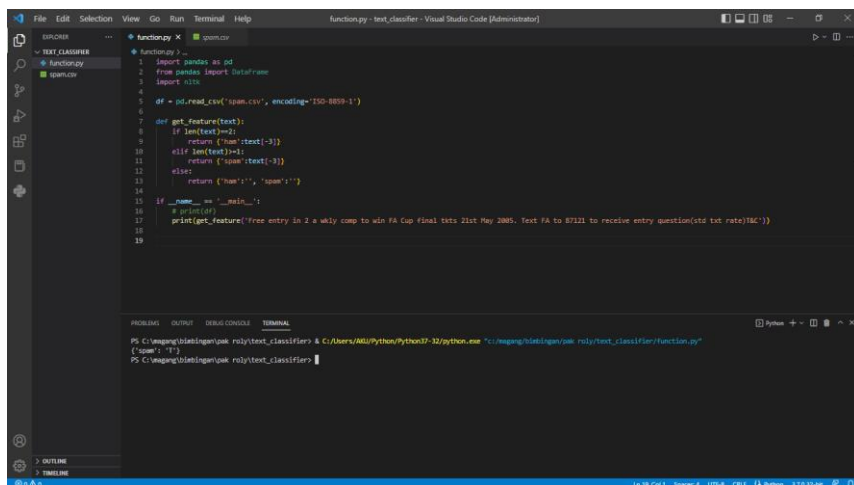
```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

v1 v2 Unnamed: 2 Unnamed: 3 Unnamed: 4
0 ham Go until jurong point, crazy.. Available only ... NaN NaN NaN
1 ham Ok lar... Joking wif u oni... NaN NaN NaN
2 spam Free entry in 2 a wkly comp to win FA Cup fina... NaN NaN NaN
3 ham U dun say so early hor... U c already then say... NaN NaN NaN
4 ham Nah I don't think he goes to usf, he lives aro... NaN NaN NaN
... ... ... ...
5567 spam This is the 2nd time we have tried 2 contact u... NaN NaN NaN
5568 ham Will i_b going to esplanade fr home? NaN NaN NaN
5569 ham Pity, * was in mood for that. So...any other s... NaN NaN NaN
5570 ham The guy did some bitching but I acted like i'd... NaN NaN NaN
5571 ham Rofl. Its true to its name NaN NaN NaN

[5572 rows x 5 columns]
PS C:\magang\bimbingan\pak roly\text_classifier>
```

Gambar 3.20 hasil output

Dan ini hasil output dari dataset nya



```
File Edit Selection View Go Run Terminal Help function.py - text_classifier - Visual Studio Code (Administrator)

TEXT_CLASSIFIER
function.py
spam.csv

1 import pandas as pd
2 from pandas import DataFrame
3 import numpy
4
5 df = pd.read_csv('spam.csv', encoding='ISO-8859-1')
6
7 def get_future(text):
8     if len(text)>2:
9         return ('ham' if text[-3])
10    elif len(text)>1:
11        return ('spam' if text[-3])
12    else:
13        return ('ham' if text[-1])
14
15 if __name__ == '__main__':
16     # print(df)
17     print(get_future('Free entry in 2 a wkly comp to win FA Cup final tickets 21st May 2005. Text FA to 87121 to receive entry question(sat txt rate)14c'))
18
19

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

PS C:\magang\bimbingan\pak roly\text_classifier> python function.py
C:\magang\bimbingan\pak roly\text_classifier
```

Gambar 3.21 hasil output get_future


```

23
24 def get_data(df, get_feature=get_feature):
25     featrues = []
26     for i, row in df.iterrows():
27         text = row['v1']; type = row['v2']
28         if isinstance(text, str):
29             if ' ' in text:
30                 text = text.replace(' ', '')
31             if '(' not in text:
32                 featrues.append((get_feature(text), type.strip('() ')))
33             else:
34                 text = text.partition('(')[0]
35                 featrues.append((get_feature(text), type.strip('() ')))
36     return featrues
37

```

Gambar 3.23 get_data

pada baris fungsi ini untuk Fungsi `isinstance()` mengembalikan True jika objek yang ditentukan adalah tipe yang diambil dari df kemudian mengganti satu karakter dari string yang diberikan lalu di tambahkan elemen ke akhir spasi dan menghapus semua karakter awal (spasi di awal) dan karakter tambahan (spasi di akhir).

```

37
38 def get_train_test(featrues, ratio=0.9):
39
40     N = len(featrues)
41     T = int(N * ratio)
42     train = featrues[:T]
43     test = featrues[T:]
44     return train, test
45

```

Gambar 3.24 get_train_test

Fungsi ini untuk membagi set data untuk melatih, ratio, menguji set data. Dataset perlu dikocok sebelum menggunakan fungsi

```
45
46 def text_classifier(df, f=get_feature):
47     data = get_data(df, f)
48     train, test = get_train_test(data)
49     classifier = nltk.NaiveBayesClassifier.train(train)
50     acc = nltk.classify.accuracy(classifier, test)
51     return classifier, acc
52
```

Gambar 3.25 text_classifier

Disini kita hanya memanggil dataset dan function get_featire lalu kita lanjutkan dan gunakan. Kami menerapkan NAIVE BAYES CLASSIFIER ke dataset kami. Kami pertama-tama membuat set data pelatihan untuk model kami, lalu kami membuat set data pengujian untuk mengujinya.

```
52
53 def show_type_of_text(text, texts=False, show_acc=False):
54     f = get_feature_text if texts else get_feature
55     classifier, acc = text_classifier(df, f)
56     if show_acc:
57         print(f'Accuracy: {acc:.4}')
58     clf = classifier.classify(f(text))
59     print(f'{text}: {clf}')
60     classifier.show_most_informative_features(10)
61
62
```

Gambar 3.26 show_type_of_text

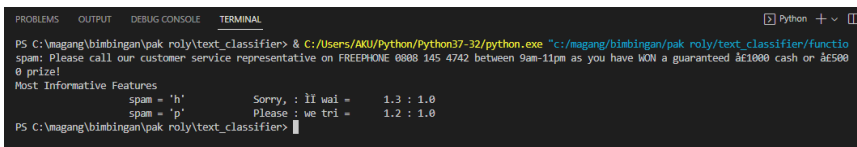
Dari hasil pemodelan dan klasifikasi yang kita buat disini ketika kita run maka akan keluar hasilnya untuk tipe spam. Kata kunci Python False mirip dengan kata kunci True, tetapi dengan nilai Boolean yang berlawanan dari false. Namun, jika kita menggunakan def show_type_of_text(text,

texts=False, show_acc=False), tipe text dan isi item(kalimat) yang ditampilkan.

```
63
64 if __name__ == '__main__':
65     # print(df)
66     # print(get_feature('Free ent
67     # print(get_feature_text('v1'
68     show_type_of_text("spam")
```

Gambar 3.27 show_type_of_text

Sekarang kita akan mencoba untuk menjalankannya



```
PS C:\magang\bimbingan\pak roly\text_classifier> & C:/Users/AGU/Python/Python37-32/python.exe "c:/magang/bimbingan/pak roly/text_classifier/functio
spam: Please call our customer service representative on FREEPHONE 0800 145 4742 between 9am-11pm as you have WON a guaranteed â€1000 cash or â€500
0 prize!
Most Informative Features
      spam = 'h'          Sorry, : If wal =      1.3 : 1.0
      spam = 'p'          Please: we tri =      1.2 : 1.0
PS C:\magang\bimbingan\pak roly\text_classifier>
```

Gambar 3.28 output

Dan ini adalah hasilnya ketika kita jalankan, maka akan terlihat tipe kalimatnya berupa spam dan isi kalimatnya juga berupa kalimat spam

```
61
62 def give_type(type1='spam', type2='ham'):
63     data = get_data(df, get_feature)
64     classifier = nltk.NaiveBayesClassifier.train(data)
65     following = classifier.prob_classify({'spam':type2, 'ham':type1})
66     x = following.generate()
67     print(f'type2: {type1}{x}')
68
```

Gambar 3.29 give_type

Pada baris ini diberikan penamaan type1 untuk spam dan type2 untuk ham pada parameter kemudian data yang di ambil dari function get_data df, dan get_feature lalu data akan di olah menggunakan library dari naive bayes lalu akan di labeling spam dan ham dan hasil nya akan di generate lalu di print.

```
59
70 if __name__ == '__main__':
71     # print(df)
72     # print(get_feature('Free entry in 2 a wkly comp to w
73     # print(get_feature_text('v1'))
74     # show_type_of_text("spam")
75     give_type(type1='Go until jurong point, crazy..')
```

Gambar 3.30 show give_type

Setelah itu coba kita masukan kalimat ham pada type1 yang berada di dataset disini saya memasukan kalimat depan nya saja.

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL
PS C:\magang\bimbingan\pak roly\text_classifier> & C:\Users\MOU\Python\Python37-32\python.exe "C:\magang\bimbingan\pak roly\text_classifier\function.py"
ham: Go until jurong point, crazy..Of course I don't tease me ... You know I simply must see I *grins* ... Do keep me posted my prey ... *loving smile* *devou
ring kiss*
PS C:\magang\bimbingan\pak roly\text_classifier> []
```

Gambar 3.31 hasil output

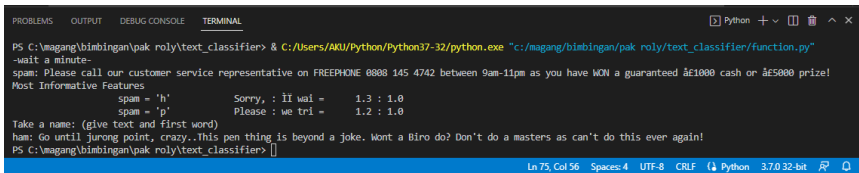
Nah ini hasilnya ketika di jalankan maka kalimat yang kita masukan di bagian depan saja maka akan keluar full pada baris kalimatnya.

```
76
77 if __name__ == '__main__':
78     print('-wait a minute-')
79     show_type_of_text("spam")
80     print('Take a name: (give text and first word)')
81     give_type(type1='Go until jurong point, crazy..')
```

Gambar 3.32 show all

Bab 3 Metodologi Penelitian

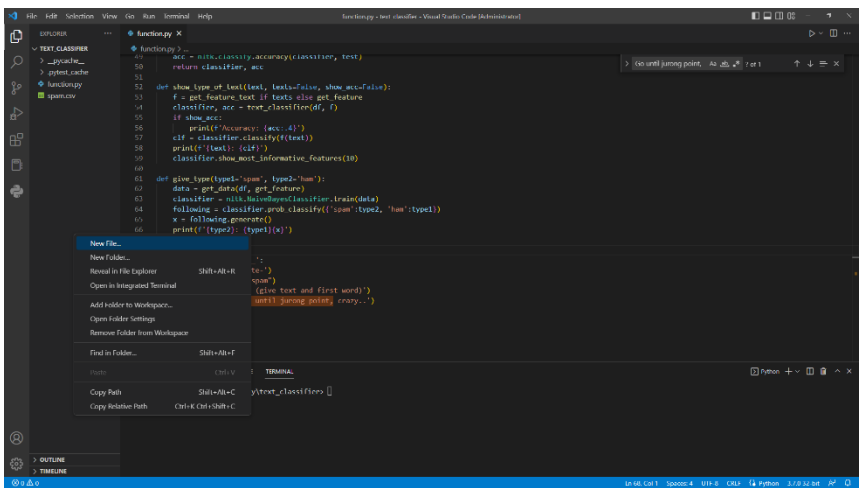
Sekarang pada bagian eksekusi ini kita ganti menjadi seperti ini untuk melihat hasil output secara keseluruhan dari program yang kita buat



```
PS C:\magang\bimbingan\pak roly\text_classifier> C:\Users\AKU\Python\Python37-32\python.exe "c:/magang/bimbingan/pak roly/text_classifier/function.py"
-wait a minute-
spam: Please call our customer service representative on FREEPHONE 8888 145 4742 between 9am-11pm as you have WON a guaranteed ££1000 cash or ££5000 prize!
Most Informative Features
spam = 'h'          Sorry, if uai =      1.3 : 1.0
spam = 'p'          Please, we tr1 =    1.2 : 1.0
Take a name: (give text and first word)
ham: Go until junong point, crazy..this pen thing is beyond a joke. Mont a Biro do? Don't do a masters as can't do this ever again!
PS C:\magang\bimbingan\pak roly\text_classifier>
```

Gambar 3.33 hasil output

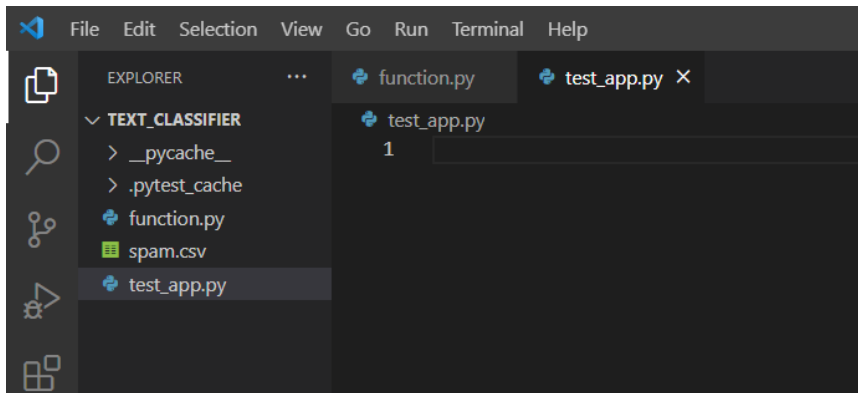
Dan ini adalah hasil output secara keseluruhan dari program yang kita buat ketika kita run



```
function.py
1: def accuracy(classifier, text):
2:     acc = classifier.accuracy(classifier, text)
3:     return acc
4:
5: def show_type_of_text(text, text2=False, show_acc=False):
6:     f = get_feature_text(text, text2)
7:     classifier, acc = test_classifier(f)
8:     if show_acc:
9:         print(f'Accuracy: {acc:.4f}')
10:    ctf = classifier.classify(f)
11:    print(f'({text}): {ctf}')
12:    classifier.show_most_informative_features(10)
13:
14: def give_type(text1='spam', text2='ham'):
15:     data = get_data(f, get_feature)
16:     classifier = nltk.MaxentClassifier.train(data)
17:     following = classifier.prob_classify(['spam', text2, 'ham', text1])
18:     w = following.generate()
19:     print(f'({text2}): {type1}[w]')
```

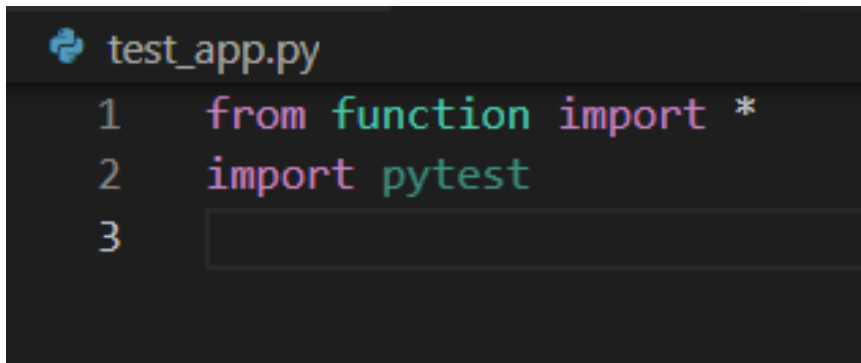
Gambar 3.34 membuat file test

Sekarang kita akan membuat file test nya dengan menambah file .py baru



Gambar 3.35 file test_app.py

Kita namakan dengan test_app.py



Gambar 3.36 import library

Pertama kita importkan terlebih dahulu dari file function dan juga import pytest

```
test_app.py > ...
1  from function import *
2  import pytest
3
4  class TestApp:
5      def test_show_type_text(self):
6          print('text:')
7          show_type_of_text("Go until jurong point, crazy..")
8          assert True
```

Gambar 3.37 test_show_type_text

Lalu disini kita buat class TestApp, lalu kita buat function sesuai dengan apa yang kita ingin test disini saya akan mengetest function dari show_type_text() dengan memasukan kalimat yang ada di dataset

```
===== test session starts =====
platform win32 -- Python 3.7.0, pytest-7.2.0, pluggy-1.0.0 -- c:\users\aku\python\python37-32\python.exe
cachedir: .pytest_cache
rootdir: C:\magang\bimbingan\pak roly\text_classifier
collected 1 item

test_app.py::TestApp::test_show_type_text PASSED [100%]

===== 1 passed in 11.99s =====
PS C:\magang\bimbingan\pak roly\text_classifier>
```

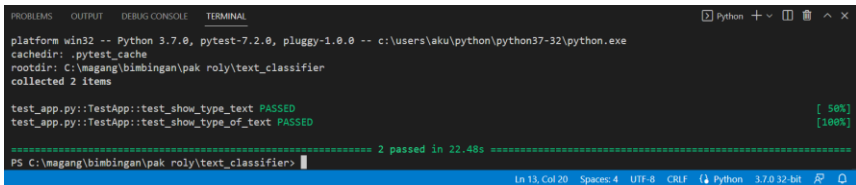
Gambar 3.38 hasil outputnya

Lalu kita run dengan pytest -v test_app.py lalu enter

```
test_app.py > TestApp > test_show_type_of_text
1  from function import *
2  import pytest
3
4  class TestApp:
5      def test_show_type_text(self):
6          print('text:')
7          show_type_of_text("Go until jurong point, crazy..")
8          assert True
9
10     def test_show_type_of_text(self):
11         print('text:')
12         show_type_of_text("FreeMsg Hey there darling it's been 3 week's now and no word back!", False)
13         assert True
```

Gambar 3.39 test_show_type_of_text

Lalu di function ini juga sama cuman bedanya untuk membaca tipe pada text nya



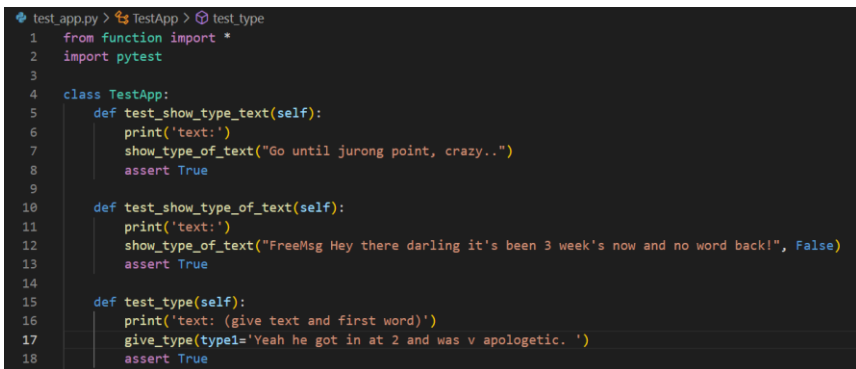
```
platform win32 -- Python 3.7.0, pytest-7.2.0, pluggy-1.0.0 -- c:\users\aku\python\python37-32\python.exe
cachedir: .pytest_cache
rootdir: C:\magang\bimbingan\pak roly\text_classifier
collected 2 items

test_app.py::TestApp::test_show_type_text PASSED [ 50%]
test_app.py::TestApp::test_show_type_of_text PASSED [100%]

===== 2 passed in 22.48s =====
PS C:\magang\bimbingan\pak roly\text_classifier>
```

Gambar 3.40 hasil outputnya

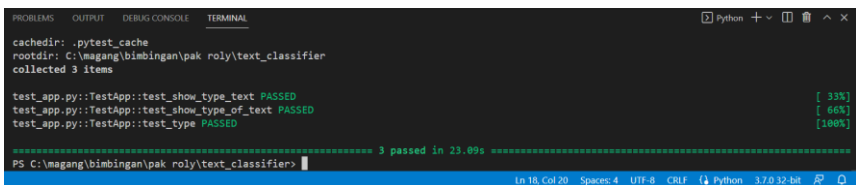
Lalu ini kita run dan ini adalah hasilnya



```
test_app.py > TestApp > test_type
1 from function import *
2 import pytest
3
4 class TestApp:
5     def test_show_type_text(self):
6         print('text:')
7         show_type_of_text("Go until jurong point, crazy..")
8         assert True
9
10    def test_show_type_of_text(self):
11        print('text:')
12        show_type_of_text("FreeMsg Hey there darling it's been 3 week's now and no word back!", False)
13        assert True
14
15    def test_type(self):
16        print('text: (give text and first word)')
17        give_type(type1='Yeah he got in at 2 and was v apologetic. ')
18        assert True
```

Gambar 3.41 test_type

Disini saya akan mengetest give_type dengan memasukan type1 dari label spam dengan kalimat yang bertipe kan ham



```
platform win32 -- Python 3.7.0, pytest-7.2.0, pluggy-1.0.0 -- c:\users\aku\python\python37-32\python.exe
cachedir: .pytest_cache
rootdir: C:\magang\bimbingan\pak roly\text_classifier
collected 3 items

test_app.py::TestApp::test_show_type_text PASSED [ 33%]
test_app.py::TestApp::test_show_type_of_text PASSED [ 66%]
test_app.py::TestApp::test_type PASSED [100%]

===== 3 passed in 23.09s =====
PS C:\magang\bimbingan\pak roly\text_classifier>
```

Gambar 3.42 hasil outputnya

Lalu kita jalankan maka akan passed

3.6 Penanganan Error

1. Error > attributeError: 'Nonetype' object has no attribute 'message'
pada proses get_data() data. Method harus diubah menjadi features ketika anda ingin ada limit proses pengambilan data dan diganti dengan itterrows().
2. Ketika ditemukan error pada salah satu cara seperti di bawah ini

```
def get_data(df, get_feature=get_feature):  
    # dataframe -> List[(feature dict, label),...]  
    featrues = []  
    for i, row in df.message: # disini masalahnya kalo di run  
        if row.startswith("E"):  
            featrues.append(i)  
            if ' ' in name:  
                name = name.replace(' ', '')  
            if '(' not in name:  
                featrues.append((get_feature(name), i.strip('(') ' '))  
            else:  
                name = na (method) append(__object: Any, /) -> None  
                featrues.append((get_feature(name), i.strip('(') ' '))  
    return featrues
```

Gambar 3.43 error 1

Salah satu solusinya adalah menggunakan cara lainnya, seperti berikut
menggunakan for in lalu pada df.message() diganti menjadi itterrows().

```
for i, row in df.iterrows():
    text = row['v1']; type = row['v2']
    if isinstance(text, str):
        if ' ' in text:
            text = text.replace(' ', '')
        if '(' not in text:
            featurues.append((get_feature(text), type.strip('(') ')))
        else:
            text = text.partition('(')[0]
            featurues.append((get_feature(text), type.strip('(') ')))
return featurues
```

Gambar 3.44 solving error

3. Karena Berfungsi untuk melakukan iterasi baris di suatu dataframe dalam bentuk pasangan (index, row).

```
if len(name)==2:
    return {'first':, 'second':name[-1]}
else:
    return {'first':DataFrame.datas(name[-2])[0], 'second':DataFrame.datas(name[-1])[0]}
>
E      AttributeError: type object 'DataFrame' has no attribute 'datas'

fungsi4.py:258: AttributeError
----- Captured stdout call -----
text:
===== short test summary info =====
FAILED test_fungsi4.py::TestApp::test_pinyin - AttributeError: type object 'DataFrame' has no attribute 'datas'
===== 1 failed, 3 passed in 16.27s =====
```

Gambar 3.45 output error 2

Kemudian pada if else yang kedua, di get_future_text akan diganti menjadi spam dan ham.

```
def get_feature_text(text):
    if len(text)==2:
        return {'spam':text[-2], 'ham':text[-1]}
    else:
        return {'spam':DataFrame.rename(text[-2])[0], 'ham':DataFrame.rename(text[-1])[0]}
```

Gambar 3.46 solving error 2

Menjadi seperti ini.


```
6 df = pd.read_csv('spam.csv', encoding='utf-8')
```

Gambar 3.47 error 3

Lalu encoding pada read_csv dan encoding nya yang salah

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL
return _read(filepath_or_buffer, kwds)
File "C:\Users\AKU\Python\Python37-32\lib\site-packages\pandas\io\parsers.py", line 454, in _read
    parser = TextFileReader(fp_or_buf, **kwds)
File "C:\Users\AKU\Python\Python37-32\lib\site-packages\pandas\io\parsers.py", line 948, in __init__
    self._make_engine(self.engine)
File "C:\Users\AKU\Python\Python37-32\lib\site-packages\pandas\io\parsers.py", line 1180, in _make_engine
    self._engine = CParserWrapper(self.f, **self.options)
File "C:\Users\AKU\Python\Python37-32\lib\site-packages\pandas\io\parsers.py", line 2010, in __init__
    self._reader = parsers.TextReader(src, **kwds)
File "pandas\libs\parsers.pyx", line 537, in pandas._libs.parsers.TextReader._cinit_
File "pandas\libs\parsers.pyx", line 711, in pandas._libs.parsers.TextReader._get_header
File "pandas\libs\parsers.pyx", line 905, in pandas._libs.parsers.TextReader._tokenize_rows
File "pandas\libs\parsers.pyx", line 2034, in pandas._libs.parsers.raise_parser_error
UnicodeDecodeError: 'utf-8' codec can't decode bytes in position 606-607: invalid continuation byte
PS C:\magang\bimbingan\pak roly\text_classifier> []
```

Gambar 3.48 output error 3

Solusinya adalah mengganti endcoding nya menjadi seperti ini ISO-8859-1

```
6 df = pd.read_csv('spam.csv', encoding='ISO-8859-1')
```

Gambar 3.49 solving error 3

Maka hasilnya akan seperti berikut ini :

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL

v1 v2 Unnamed: 2 Unnamed: 3 Unnamed: 4
0 ham Go until jurong point, crazy.. Available only ... NaN NaN NaN
1 ham Ok lar... Joking wif u oni... NaN NaN NaN
2 spam Free entry in 2 a wkly comp to win FA Cup fina... NaN NaN NaN
3 ham U dun say so early hor... U c already then say... NaN NaN NaN
4 ham Nah I don't think he goes to usf, he lives aro... NaN NaN NaN
... .. NaN NaN NaN
5567 spam This is the 2nd time we have tried 2 contact u... NaN NaN NaN
5568 ham Will i_b going to esplanade fr home? NaN NaN NaN
5569 ham Pity, * was in mood for that. So...any other s... NaN NaN NaN
5570 ham The guy did some bitching but I acted like i'd... NaN NaN NaN
5571 ham Rofl. Its true to its name NaN NaN NaN

[5572 rows x 5 columns]
PS C:\magang\bimbingan\pak roly\text_classifier> 
```

Gambar 3.50 Hasil

Bab IV

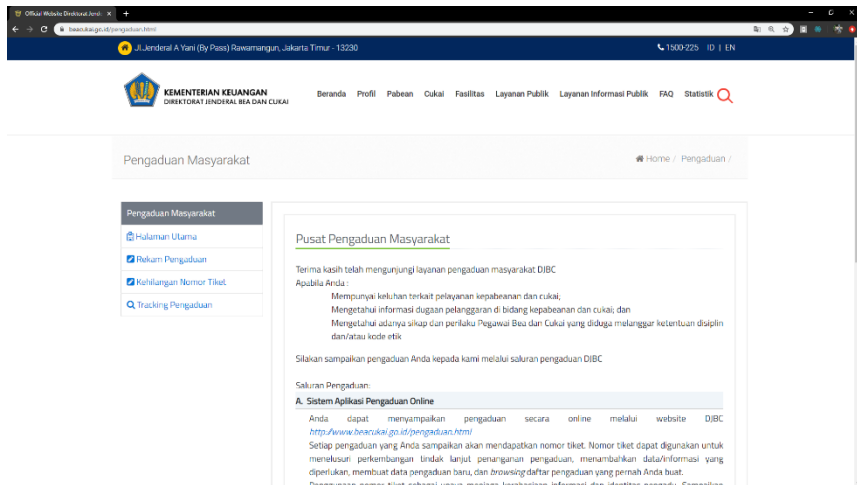
Analisis dan Hasil Pembahasan

4.1 Analisis dan Perancangan Sistem

Analisis sistem adalah penguraian sistem informasi secara utuh ke dalam beberapa bagian sistem informasi yang dirancang untuk memberikan penjabaran dan memberikan penilaian terhadap implementasi sistem informasi tersebut, yang dapat dilihat dari berbagai sisi yaitu masalah, peluang, hambatan yang terjadi, serta kebutuhan sistem informasi yang dirancang agar dapat dilakukan Perbaikan untuk memenuhi kebutuhan pengguna. Pada bab ini akan dipaparkan tentang prosedur analisis sistem aplikasi yang dirancang berupa flowmap, coding, dan analisis sistem non-fungsional yang terdiri dari perangkat keras dan perangkat lunak serta analisis pengguna yang terlibat dalam penggunaan SMS Spam Identification Sistem menggunakan Metode Multinomial Naïve Bayes.

4.1.1 Analisis Sistem Berjalan (*Current System*)

Analisis sistem yang sedang berjalan terdiri dari analisis sistem manual yang terdiri dari proses pengaduan konten melalui website Beacukai.go.id Provinsi Jakarta Timur.



Gambar 4.1 Website Bea Cukai Pengaduan

4.1.1.1 Analisis Prosedur yang Berjalan (*Flowmap/Flowchart*)

Dalam prosedur ini dilibatkan 3 entitas (aktor) yaitu pelapor/user, admin, dan sistem. Pelapor membuka website <https://www.beacukai.go.id/> kemudian memilih opsi Pengaduan Masyarakat > Rekaman Pengaduan.

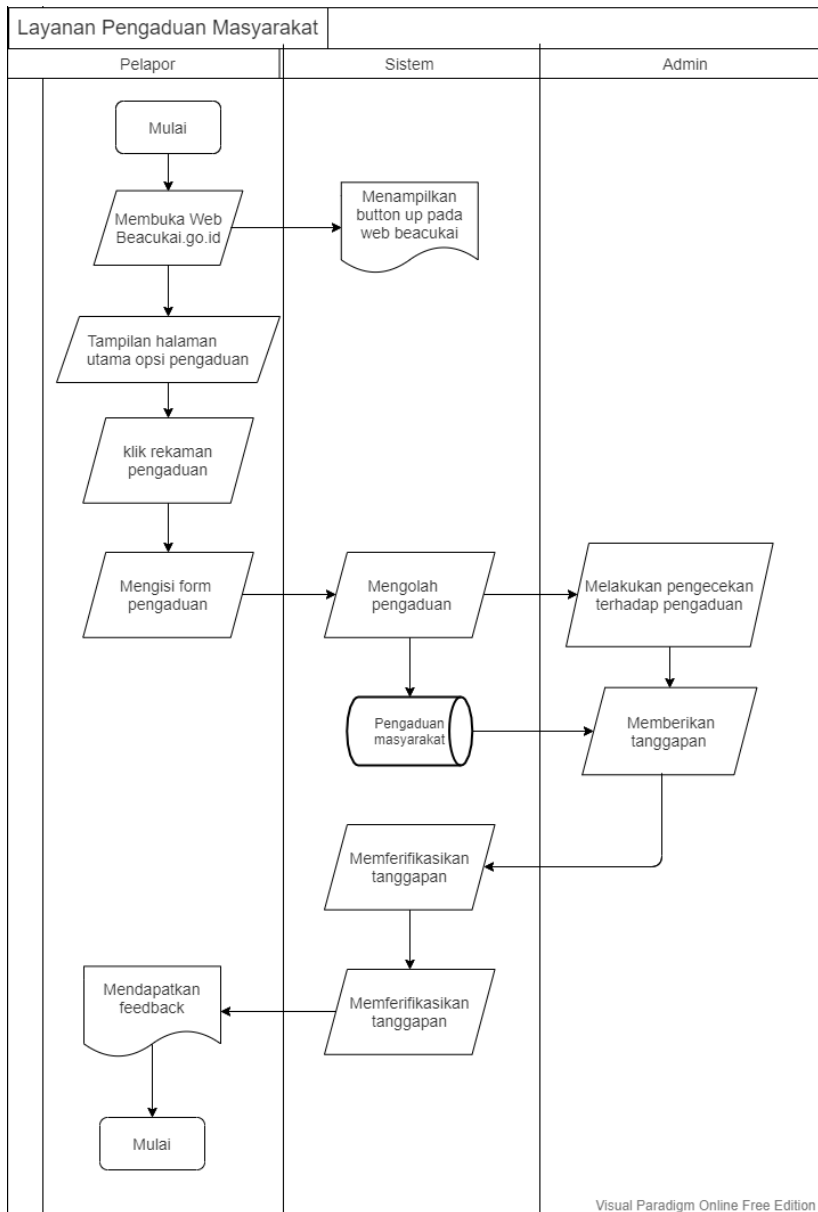


Gambar 4.2 Opsi Pengaduan

Kemudian mengisi form pengaduan dan laporkan keresahan yang terjadi di bidang IKC.

Gambar 4.3 Form Pertanyaan / Aduan

Adapun flowmap prosedur Layanan Pengaduan Konten sebagai berikut:



Gambar 4.4 Flowmap Prosedur Layanan Pengaduan

Penjelasan:

1. User/pelapor membuka halaman website Beacukai.go.id
2. Sistem menampilkan halaman utama website Beacukai
3. User/pelapor melihat halaman kemudian memilih menu Pengaduan Masyarakat dan opsi rekam pengaduan
4. Sistem menampilkan form halaman pengaduan
5. Sistem mengolah aduan yang diajukan oleh user/pelapor kemudian menyimpannya di database
6. Admin melakukan pengecekan terhadap aduan
7. Admin memberikan tanggapan atas aduan yang diajukan
8. Sistem memverifikasi tanggapan dari admin dan memberikan feedback terhadap user/pelapor
9. User/pelapor mendapatkan feedback

4.2 Perancangan Arsitektur Perangkat Lunak dan Perangkat Keras Sistem

Analisis kebutuhan non fungsional dilakukan untuk mengetahui spesifikasi kebutuhan sistem. Spesifikasi kebutuhan melibatkan analisis perangkat keras/*hardware*, analisis perangkat lunak/*software*, analisis pengguna/*user*.

4.2.1 Kebutuhan Perangkat Keras

Tabel 4.1 Deskripsi Perangkat Keras Server

No.	Nama Perangkat	Spesifikasi	Keterangan
1.	Harddisk	-	Media untuk menyimpan data aplikasi yang dibuat
2.	Memory	-	Memori sistem yang digunakan
3.	Processor	-	Untuk kecepatan transfer data dari sistem yang sangat bergantung pada kecepatan prosesor komputer
4	PC (<i>Personal Computer</i>)	-	Perangkat utama dalam jaringan komputer yang mengirim dan mengakses data dalam jaringan
5	NIC	ISA dan PCI	<i>Port</i> yang menghubungkan komputer dengan jaringan
6	Infrastruktur jaringan		<i>Server</i> dan <i>host</i>

Tabel 4.2 Deskripsi Perangkat Keras Klien

No.	Nama Perangkat	Spesifikasi	Keterangan
1	<i>Harddisk</i>	-	Sebagai tempat untuk menyimpan data yang dibutuhkan tetapi pada sisi <i>client</i> tidak diharuskan memiliki ketersediaan <i>space</i> yang besar
2	<i>Memory</i>	-	Kecepatan <i>client</i> dalam mengakses <i>system</i> ini
3	Infrastruktur Jaringan	-	<i>Server</i> dan <i>Host</i>
4	PC (<i>Personal Computer</i>) atau Gadget (<i>Mobile device</i>)	-	Perangkat utama dalam jaringan komputer yang mengirim dan mengakses data dalam jaringan

4.2.2 Kebutuhan Perangkat Lunak

Tabel 4.3 Deskripsi Perangkat Lunak Server

No.	Tools/Software	Fungsi	Keterangan
1.	Windows	Sistem Operasi	-
2.	Python, HTML, Bootstrap, Flask	Bahasa Pemrograman yang digunakan	-
3.	Web Server	Browser	-
4.	Visual Studio Code	Interface	-

Tabel 4. Deskripsi Perangkat Lunak Client

No.	Tools/Software	Fungsi	Keterangan
1.	<i>Windows, Android</i>	Sistem Operasi	-
2.	<i>Chrome, Firefox, Opera, Chrome, dst.</i>	<i>Web Browser</i>	-

Bab V

Penutup

5.1 Kesimpulan

Analisis data merupakan topik yang sedang hangat dibicarakan, terutama oleh teman-teman yang mendalami dunia teknologi. Data dapat dimanfaatkan sedemikian rupa untuk pembelajaran mesin dan diimplementasikan ke dalam sistem yang sederhana.

Berdasarkan analisis data yang telah dilakukan, diperoleh beberapa kesimpulan sebagai berikut, antara lain; data diperoleh dengan crawling data kaggle, preprocessing data yang terdiri dari tokenization, filtering stopwords, training, dan testing untuk mengubah data mentah menjadi data yang lebih bermakna sehingga dapat dipahami oleh algoritma yang akan digunakan.

Algoritma yang digunakan untuk melatih data dan memprediksi adalah algoritma Multinomial Naïve Bayes yang merupakan salah satu dari banyak jenis algoritma untuk pengklasifikasi khususnya klasifikasi teks. Setelah melakukan uji coba data training dan testing didapatkan akurasi skor sebesar 1.0. Buku ini disusun dengan maksud dan tujuan untuk memberikan wawasan dan sebagai sumber referensi bacaan mengenai analisis data dengan menggunakan algoritma Multinomial Naïve Bayes. Melalui buku ini, pembaca diharapkan mengetahui betapa pentingnya data, bagaimana

pemanfaatan data, dan apa saja yang harus dilakukan saat menganalisis data, langkah apa saja yang perlu dilakukan sebelum akhirnya diimplementasikan ke dalam sistem yang sederhana.

Daftar Pustaka

- Abdulloh, Rohi. (2015) Web Programing is Easy. Jakarta: PT Elex Media Komputindo.
- Arhami, Muhammad. (2005) Konsep Dasar Sistem Pakar. Yogyakarta: Andi Offset.
- Arief M Rudianto. (2011) Pemrograman Web Dinamis menggunakan PHP dan MySQL. Yogyakarta: Andi Offset.
- Andre. (2013, oktober 01). Pengertian CSS, Apa yang dimaksud dengan CSS. Retrieved 06 12, 2020, from <https://www.duniaikom.com/tutorial-belajar-css-part-1-pengertian-css-apa-yang-dimaksud-dengan-css/>
- Barber, I. 2010. Bayesian Opinion Mining. [Online]. Tersedia di: <http://phpir.com/bayesian-opinion-mining> [diunduh: 15 Juni 2014].
- Feldman, R & Dagan, I. (1995) Knowledge discovery in textual databases (KDT). Dalam Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95), Montreal, Canada, August 20-21, AAAI Press, 112-117.
- Feldman, R & Sanger, J. (2007) The Text Mining Handbook-Advanced Approaches in Analyzing Unstructured Data, USA: New York.
- Go, Alec; Bhayani, Richa; Huang, Lei. (2009) Twitter Sentimen Classification using Distant Supervision.

- H, Azizul, “Berkenalan dengan scikit-learn (Part 1) – Preparations,” hkaLabs, 2018. [Online]. Available: <https://hakim-azizul.com/berkenalan-dengan-scikitlearn/>.
- H, Mooney. (2018) “Fake news” and the Sociological Imagination: Theory Informs Practice”. LOEX Quarterly, Vol.44 No.4: 4-16.
- Hall, J, A., (2004) Sistem Informasi Akuntansi. Jakarta: Penerbit Salemba Empat.
- Hearst, M. A. (1997) Text data mining: Issues, techniques, and the relationship to information access. Presentation notes for UW/MS workshop on data mining, July 1997.
- Hidayatullah, dan Kawistara. (2015) Pemrograman Web. Bandung: Informatika.
- Hoffer. (2002) dalam Laporan Kerja Praktik 2 Sistem Monitoring Pendistribusian Surat oleh Dwi Oktarina Sari. (2006).
- Ilham, M. (2020) Pengertian Python, Fungsi, Kelebihan dan Kekurangan. Retrieved from [materibelajar.co.id: https://materibelajar.co.id/pengertianpython/](https://materibelajar.co.id/pengertianpython/)
- Jhonsen. (2004) WEB DESIGNER untuk PEMULA. Jakarta : Elex Media Komputindo Kelompok Gramedia. Anggota IKAPI.
- Jogiyanto. (2008) Metodologi Penelitian Sistem Informasi. Yogyakarta: Andi Offset.
- K.A. Wibowo, D. Rahmawan and E. Maryani, (2019) In Indonesia, young and old share fake news on social media, The Conversation.

- Kusrini. (2008) Aplikasi Sistem Pakar. Yogyakarta: Andi Offset.
- La Midjan dan Susanto, A,. (2008) Sistem Informasi Akuntansi Konsep dan Pengembangannya. Bandung: Penerbit Lingga Jaya.
- Lamurias, Andre Francisco.M. 2014. Identifying Interactions Between Chemical Entities in Text. Desertation Universidade De Lisboa.
- M. K. Albert Verasius Dian Sano, S.T., “DEFINISI, KARAKTERISTIK, DAN MANFAAT DATA MINING -SERI DATA MINING FOR BUSINESS INTELLIGENCE (2),” BINUS UNIVERSITY, 2019. [Online]. Available: <https://binus.ac.id/malang/2019/01/definisi-karakteristikdan-manfaat-data-mining-seri-data-mining-for-businessintelligence-2/>.
- McLeon. (2008) Sistem Informasi Manajemen. Jakarta: Salemba Empat.
- Mudjahidin, & Putra, N. P. (2010) Rancang Bangun Sistem Informasi Monitoring Perkembangan Proyek Berbasis Web Studi Kasus Di Dinas Bina Marga dan Pemantusan. Jurnal Teknik Industri Vol.11 No.1, 75- 83.
- R. Irsyad, “Penggunaan Python Web Framework Flask Untuk Pemula,” 2018.
- Rio, (2005) Pengertian HTML . Bogor: Universitas Bogor Indah.
- Romney, M., Steinbart, P, J,. (2005) Accounting Information System 9th Edition. Jakarta: Penerbit Salemba Empat.
- S. Kumar and N. Shah, False Information on Web and Social Media: A Survey. 1, 1 (April 2018), <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>.

- Saputra, A. (2014) API Developer Buku Sakti Para Pengembang Web.
- Sutabri, T. (2012) Konsep Sistem Informasi. Yogyakarta : C.V. Andi Offset.
- Syamsul. (2003) Manajemen Operasi. Bogor: Grasindo.
- Tarlach McGonagle, Fake News: False Fears or Real Concerns?,
Netherlands Quartely of Human Rights, 2017, hlm. 203.
- Triawati, Candra; Bijaksana, M.Arif; Indrawati, Nur; Saputro, Widyanto Adi.
(2009) Pemodelan Berbasis Konsep Untuk Kategorisasi Artikel Berita
Berbahasa Indonesia. Dalam Seminar Nasional Aplikasi Teknologi
Informasi 2009
- Twitter, 04 Maret (2016), Twitter API,
<https://dev.twitter.com/overview/documentation>
- V.L. Rubin, Y. Chen and N.J. Conroy, (2015) Deception detection for news:
three types of fakes. Proceedings of the 78th ASIS&T Annual
Meeting: Information Science with Impact: Research in and for the
Community, American Society for Information Science, 83.
- V.L., Rubin. (2016) Deception detection and rumor debunking for social
media, The SAGE Handbook of Social Media Research Methods:
342-364.
- William P. Wagner. (2017) Trends in Expert System Development : A
Longitudinal Content Analysis of Over Thirty Years of Expert System
Case Studies. Expert System With Applications, 85-96.

Lampiran

1. Lampiran data Text_SMS.csv

```
      v1      v2 Unnamed: 2 Unnamed: 3 Unnamed: 4
0      ham  Go until jurong point, crazy.. Available only ...      NaN      NaN      NaN
1      ham                Ok lar... Joking wif u oni...      NaN      NaN      NaN
2      spam  Free entry in 2 a wkly comp to win FA Cup fina...      NaN      NaN      NaN
3      ham  U dun say so early hor... U c already then say...      NaN      NaN      NaN
4      ham  Nah I don't think he goes to usf, he lives aro...      NaN      NaN      NaN
...      ...      ...      ...      ...
5567 spam  This is the 2nd time we have tried 2 contact u...      NaN      NaN      NaN
5568 ham                Will i_ b going to esplanade fr home?      NaN      NaN      NaN
5569 ham  Pity, * was in mood for that. So...any other s...      NaN      NaN      NaN
5570 ham  The guy did some bitching but I acted like i'd...      NaN      NaN      NaN
5571 ham                Rofl. Its true to its name      NaN      NaN      NaN

[5572 rows x 5 columns]
```

Gambar Lampiran - Data Text SMS