

**ANALISIS SENTIMEN ULASAN PRODUK PADA YOUTUBE
MENGUNAKAN METODE NAÏVE BAYES**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



diajukan oleh

AYUB MASRUKAN

18.11.1849

Kepada

PROGRAM SARJANA

PROGRAM STUDI S1 INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2022

**ANALISIS SENTIMEN ULASAN PRODUK PADA YOUTUBE
MENGUNAKAN METODE NAÏVE BAYES**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



diajukan oleh
AYUB MASRUKAN
18.11.1849

Kepada
PROGRAM SARJANA
PROGRAM STUDI S1 INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022

HALAMAN PERSETUJUAN
SKRIPSI
ANALISIS SENTIMEN ULASAN PRODUK PADA YOUTUBE
MENGGUNAKAN METODE NAÏVE BAYES

yang disusun dan diajukan oleh

Ayub Masrukan

18.11.1849

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Agustus 2022

Dosen Pembimbing,

Windha Mega Pradnya D, M.Kom
NIK. 190302185

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN ULASAN PRODUK YOUTUBE MENGGUNAKAN
METODE NAÏVE BAYES**

yang disusun dan diajukan oleh

Ayub Masrukan

18.11.1849

Telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Agustus 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Mardhiya Hayaty, S.T., M.Kom.
NIK. 190302108

Ike Verawati, M.Kom
NIK. 190302237

Windha Mega Pradnya D, M.Kom
NIK. 190302185

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Agustus 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta,S.Kom., M.Kom.
NIK. 19030209

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Ayub Masrukan
NIM : 18.11.1849

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen Ulasan Produk Pada Youtube Menggunakan Metode Naïve Bayes

Dosen Pembimbing : Windha Mega Pradnya D, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 31 Agustus 2022

Yang Menyatakan,



Ayub Masrukan

MOTTO

“Jalani, Nikmati, Syukuri”

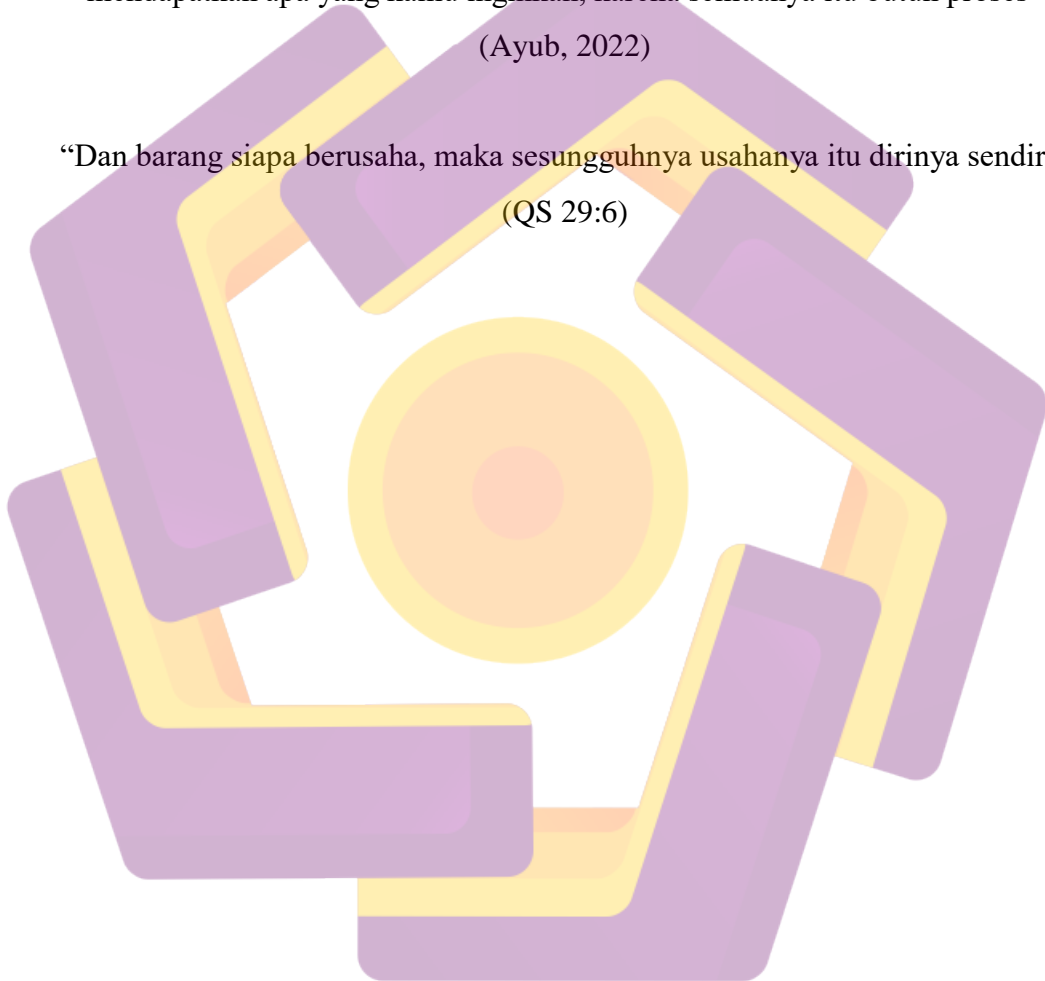
(Ayub, 2020)

“Kerjakanlah semuanya dengan sungguh-sungguh dan kamu pasti akan mendapatkan apa yang kamu inginkan, karena semuanya itu butuh proses”

(Ayub, 2022)

“Dan barang siapa berusaha, maka sesungguhnya usahanya itu dirinya sendiri”

(QS 29:6)



HALAMAN PERSEMBAHAN

Alhamdulillah, saya berhasil menyelesaikan skripsi ini dengan maksimal. Hal ini tentunya tak terlepas dari karunia, rahmat serta hidayah yang telah diberikan oleh Allah swt. Sehingga saya mendapatkan kemudahan, kelancaran dan kekuatan dalam mengerjakan skripsi ini. Selain itu ada orang-orang hebat di belakang saya yang selalu memberikan dukungan dan motivasinya baik secara langsung maupun tidak langsung, antara lain.

1. Orang tua serta keluarga besar saya yang selalu memberikan motivasi, restu dan doa tanpa henti.
2. Ibu Windha Mega Pradya D., M.Kom., yang telah membimbing saya dalam proses mengerjakan skripsi ini.
3. Fahry, M.Kom yang menjelma sebagai dosen pembimbing dikala saya mengalami *stuck* dalam proses mengerjakan skripsi ini dan juga Alfian serta Ronald.
4. Teman-teman Pengurus Inti HMIF 2020/2021 yang terus berjuang menemani saya disaat suka maupun duka, disaat susah maupun bahagia sampai akhir kepengurusan.
5. Teman-teman dari Himpunan Mahasiswa Informatika yang sudah saya anggap sebagai keluarga saya ketika di Jogja.
6. Teman-teman seperjuangan dari kelas 18-IF-01 yang telah menemani saya dari semester awal hingga akhir, terutama yang sering saya repotkan dalam mengerjakan tugas kuliah.

KATA PENGANTAR

Puji syukur kehadiran Allah swt. yang telah memberikan karunia, rahmat serta hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik dan maksimal. Skripsi yang berjudul **“Analisis Sentimen Ulasan Produk Pada Youtube Menggunakan Metode Naïve Bayes”** ini disusun sebagai salah satu syarat dalam menyelesaikan masa studi program sarjana di Universitas AMIKOM Yogyakarta.

Dalam penulisan skripsi ini, penulis mengucapkan rasa terimakasih atas motivasi, bimbingan, saran dan masukan dari berbagai pihak secara moral maupun spiritual. Pada kesempatan kali ini penulis mengucapkan terimakasih kepada :

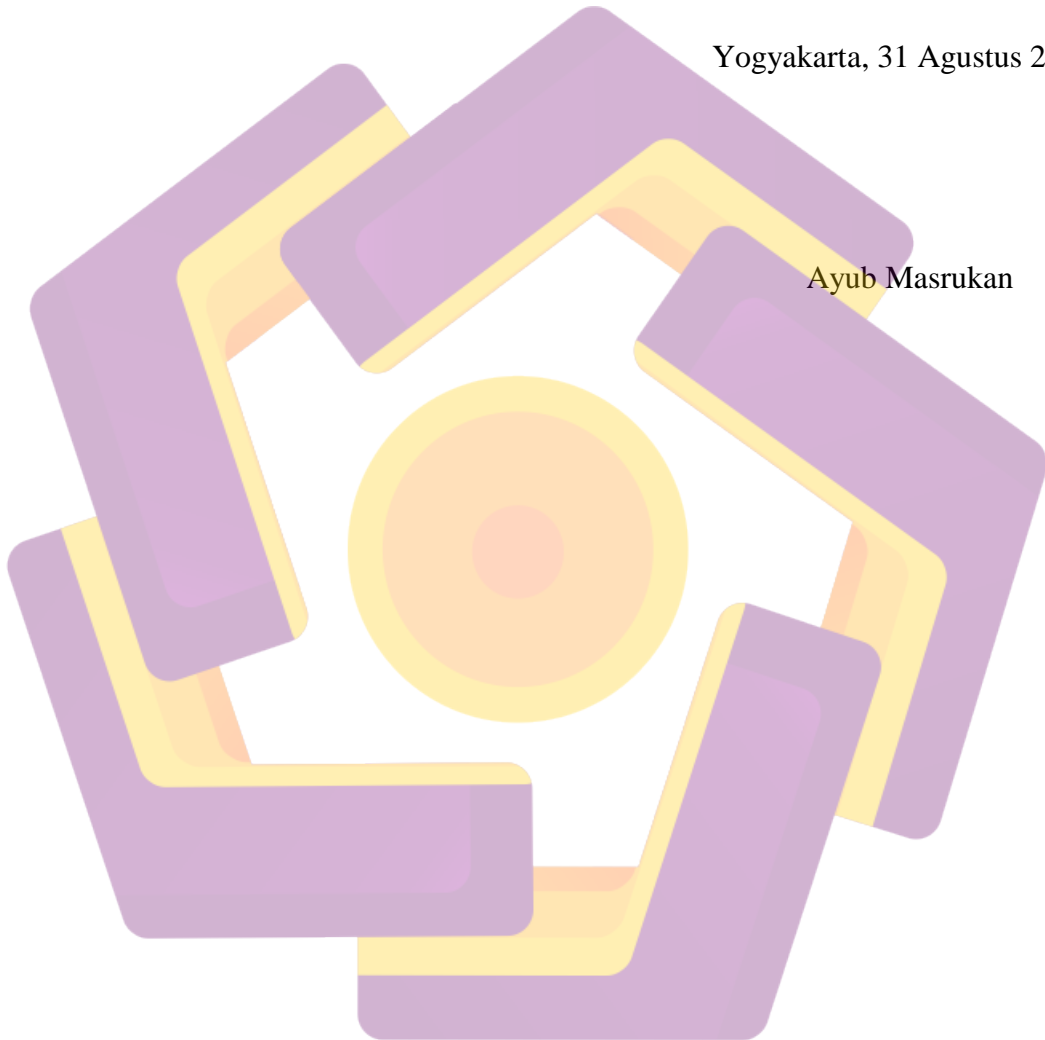
1. Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, M. Kom selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Windha Mega Pradya D., M.Kom selaku Ketua Program Studi S1 Informatika Universitas AMIKOM Yogyakarta dan selaku Dosen Pembimbing yang selalu memberikan bimbingan, saran dan masukan dalam proses penulisan skripsi.
4. Ibu Mardhiya Hayaty, S.T., M.Kom dan Ibu Ike Verawati, M.Kom selaku Dosen Penguji yang telah memberikan saran dan masukan agar penelitian ini menjadi lebih baik.

Penulis menyadari skripsi ini masih terdapat banyak kekurangan, maka dari itu kritik dan saran membangun sangat penulis butuhkan untuk perbaikan dan kesempurnaan karya selanjutnya. Semoga Allah swt. selalu melimpahkan rahmat dan maghrih-Nya kepada kita semua.

Akhir kata, semoga penyusunan skripsi ini dapat bermanfaat bagi pembaca dalam menambah wawasan dan pengetahuan khususnya dalam bidang Informatika.

Yogyakarta, 31 Agustus 2022

Ayub Masrukan



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
MOTTO	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
INTISARI	xv
Abstract.....	xvi
BAB I.....	1
1.1 Latar Belakang	1
1.2 Rumusan masalah.....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Metodologi Penelitian	4
1.6.1 Metode Pengumpulan Data	4
1.6.2 Metode Analisis Data	5
1.6.3 Metode Implementasi	5
1.6.4 Metode Pengujian.....	5

1.7	Sistematika Penulisan.....	6
BAB II	7
2.1	Kajian Pustaka.....	7
2.2	Analisis Sentimen.....	13
2.3	Youtube	13
2.4	<i>Scrapping</i>	14
2.5	<i>Text Mining</i>	14
2.6	<i>Preprocessing</i>	15
2.7	<i>Sentistrength</i>	15
2.8	Klasifikasi.....	16
2.9	Machine Learning.....	16
2.10	<i>Naïve Bayes Classifier</i>	17
2.11	<i>K-fold Cross Validation</i>	18
2.12	<i>Confusion Matrix</i>	19
2.13	<i>Laravel</i>	20
2.14	<i>Python</i>	21
BAB III	22
3.1	Gambaran Umum	22
3.2	Tahapan Penelitian	22
3.3	Alat dan Bahan Penelitian	23
3.3.1	Alat Penelitian	24
3.3.2	Bahan Penelitian.....	24
3.3.5	Preprocessing	31
3.4	Perancangan <i>User Interface</i>	43
BAB IV	45

4.1	Persiapan dan Pengumpulan Data	45
4.2	<i>Search</i>	45
4.3	<i>Dataset</i>	45
4.4	Preprocessing	46
4.4.1	<i>Case Folding</i>	46
4.4.2	<i>Tokenisasi</i>	48
4.4.3	<i>Normalization</i>	49
4.4.4	<i>Stemming</i>	51
4.5	Proses Klasifikasi Naïve Bayes	53
4.6	Mengambil <i>Dataset</i> Berdasarkan Label.....	59
4.7	Menghitung Data Berdasarkan <i>Fold</i>	60
4.8	Mengambil <i>Data Training</i> Berdasarkan <i>Fold</i>	61
4.9	Mengambil <i>Data Testing</i> Berdasarkan <i>Fold</i>	61
4.10	Menghitung Probabilitas	62
4.11	Memberikan Label Pada Hasil Probabilitas	63
4.12	Membagi Data Dengan <i>K-Fold = 5</i>	63
4.13	Menghitung Tingkat Akurasi Dengan <i>Confusion Matrix</i>	64
4.14	Fitur <i>User Interface</i>	66
4.15	Klasifikasi <i>Naïve Bayes</i>	72
4.16	Hasil Akurasi.....	73
BAB V	74
5.1	Kesimpulan.....	74
5.2	Saran.....	74
DAFTAR PUSTAKA	75

DAFTAR TABEL

Tabel 2.1 Hasil Review Penelitian	9
Tabel 2.2 Algoritma SentiStrength	16
Tabel 2.4 K-Fold Cross Validaton dengan K Sebesar 3	19
Tabel 2. 5Confusion Matrix	19
Tabel 3.1 Spesifikasi Perangkat Keras	24
Tabel 3.2 Spesifikasi Perangkat Lunak	24
Tabel 3.3 Data Setelah Memasukkan <i>Keyword</i>	26
Tabel 3.4 Contoh <i>Dataset</i>	29
Tabel 3.5 Contoh Case <i>Folding</i>	31
Tabel 3.6 Contoh <i>Tokenisasi</i>	33
Tabel 3.7 Contoh <i>Normalization</i>	35
Tabel 3.8 Contoh <i>Stemming</i>	36
Tabel 3.9 Contoh <i>Sentistrength</i>	37
Tabel 4.1 Data Uji	53
Tabel 4.2 Data Uji	53
Tabel 4.3 Data Uji	53
Tabel 4.4 Nilai Probablilitas Tiap Kata Pada Tiap Kelas	54
Tabel 4.5 <i>K-Fold Validation</i>	64
Tabel 4.6 <i>Confusion Matrix</i>	72
Tabel 4.7 Akurasi, Presisi, <i>Recall</i>	72
Tabel 4.8 Hasil Akurasi	73

DAFTAR GAMBAR

Gambar 3.1 Diagram Alur Tahapan Penelitian.....	23
Gambar 3.2 Diagram Memasukkan <i>Keyword</i>	26
Gambar 3.3 Diagram Pengambilan <i>Dataset</i>	29
Gambar 3.4 Halaman <i>Login</i>	44
Gambar 3.5 Halaman <i>Dashboard</i>	44
Gambar 4.1 Mendapatkan <i>Channel Youtube</i>	45
Gambar 4.2 Mendapatkan <i>Dataset</i>	46
Gambar 4.3 <i>Syntax Case Folding</i>	47
Gambar 4.4 <i>Syntax Case Folding</i>	47
Gambar 4.5 <i>Syntax Case Folding</i>	48
Gambar 4.6 <i>Syntax Tokenisasi</i>	48
Gambar 4.7 <i>Syntax Tokenisasi</i>	49
Gambar 4.8 <i>Syntax Tokenisasi</i>	49
Gambar 4.9 <i>Syntax Normalization</i>	50
Gambar 4.10 <i>Syntax Normalization</i>	50
Gambar 4.11 <i>Syntax Normalization</i>	51
Gambar 4.12 <i>Syntax Stemming</i>	52
Gambar 4.13 <i>Syntax Stemming</i>	52
Gambar 4.14 <i>Syntax Stemming</i>	52
Gambar 4.15 <i>Syntax Naïve Bayes</i>	57
Gambar 4.16 <i>Syntax Naïve Bayes</i>	58
Gambar 4.17 <i>Syntax Naïve Bayes</i>	58
Gambar 4.18 <i>Syntax Naïve Bayes</i>	59
Gambar 4.19 <i>Syntax Naïve Bayes</i>	59
Gambar 4.20 <i>Syntax Mengambil Dataset Berdasarkan Label</i>	59
Gambar 4.21 <i>Syntax Menghitung Data Berdasarkan Fold</i>	60
Gambar 4.22 <i>Syntax Menghitung Data Berdasarkan Fold</i>	61
Gambar 4.23 <i>Syntax Data Training</i>	61
Gambar 4.24 <i>Syntax Data Testing</i>	61

Gambar 4.25 Menghitung Probabilitas	62
Gambar 4.26 Menghitung Probabilitas	63
Gambar 4.27 Memberikan Label Pada Probabilitas	63
Gambar 4.28 Tingkat Akurasi Dengan <i>Confusion Matrix</i>	65
Gambar 4.29 Tingkat Akurasi Dengan <i>Confusion Matrix</i>	65
Gambar 4.30 Tingkat Akurasi Dengan <i>Confusion Matrix</i>	66
Gambar 4.31 <i>User Interface Login</i>	67
Gambar 4.32 <i>User Interface Home</i>	67
Gambar 4.33 <i>User Interface Search</i>	67
Gambar 4.34 <i>User Interface Dataset</i>	68
Gambar 4.35 <i>User Interface Dataset</i>	68
Gambar 4.36 <i>User Interface Dataset</i>	69
Gambar 4.37 <i>User Interface Dataset</i>	69
Gambar 4.38 <i>User Interface Sentisstrength</i>	70
Gambar 4.39 <i>User Interface Naïve Bayes</i>	71
Gambar 4.40 <i>User Interface Naïve Bayes</i>	71
Gambar 4.41 Hasil <i>Naïve Bayes K-Fold</i>	72

INTISARI

Dengan perkembangan dunia internet yang semakin pesat di Indonesia, tentunya membuat masyarakat Indonesia tidak dapat lepas dari media sosial dalam kehidupan sehari-hari. YouTube merupakan situs web *sharing* nomer satu di dunia yang tentunya banyak konten-konten yang disajikan oleh pembuat konten YouTube atau yang biasa disebut dengan konten *creator*[1]. Berbagai sumber dapat diakses kapanpun tanpa terbatas oleh jarak. Selain itu konten atau pesan yang disampaikan dapat diterima oleh orang lain secara cepat.

Sesuai namanya, *machine learning* mencoba menirukan bagaimana proses manusia atau makhluk *cerdas* belajar dan menggeneralisasi. *Naïve Bayes Classifier* merupakan salah satu teknik klasifikasi metode probabilistik dan statistik yang ditemukan oleh Thomas Bayes yang dikenal dengan *Teorema Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya.

Dari hasil pengujian *f1-score* dengan menetapkan $k = 5$, pada *fold* 1 mendapatkan *f1-score* 64%, pada *fold* 2 mendapatkan *f1-score* 65%, pada *fold* 3 mendapatkan *f1-score* 64%, pada *fold* 4 mendapatkan *f1-score* 64%, dan pada *fold* 5 mendapatkan *f1-score* 66%.

Kata kunci: analisis sentimen, naïve bayes, komentar youtube, *f1-score*, ulasan produk.

Abstract

With the rapid development of the internet world in Indonesia, of course, Indonesian people cannot be separated from social media in their daily lives. YouTube is the number one sharing website in the world, of course a lot of content is presented by YouTube content creators or commonly called content creators [1]. Various sources can be accessed at any time without being limited by distance. In addition, the content or message conveyed can be received by others quickly.

As the name implies, machine learning tries to imitate how humans or intelligent creatures learn and generalize. Naïve Bayes Classifier is one of the probabilistic and statistical method classification techniques invented by Thomas Bayes known as the Bayes theorem, which predicts future opportunities based on previous experience.

From the results of the f1-score test by setting $k = 5$, in fold 1 you get an f1-score of 64%, in fold 2 you get an f1-score of 65%, in fold 3 you get an f1-score of 64%, in fold 4 you get an f1-score of 64 %, and on fold 5 get an f1-score of 66%.

Keyword: *sentiment analysis, nave bayes, youtube comments, f1-score, product reviews.*