

**IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK KLASIFIKASI
KOMENTAR SPAM PADA INSTAGRAM**

SKRIPSI



disusun oleh
Beta Priyoko
15.11.8912

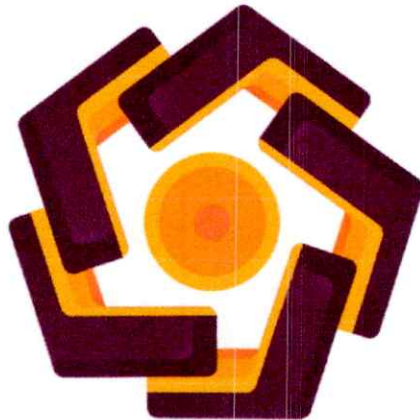
**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2019**



**IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK KLASIFIKASI
KOMENTAR SPAM PADA INSTAGRAM**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh
Beta Priyoko
15.11.8912

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2019



PERSETUJUAN

SKRIPSI

IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK KLASIFIKASI KOMENTAR SPAM PADA INSTAGRAM

yang dipersiapkan dan disusun oleh

Beta Priyoko

15.11.8912

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 25 Februari 2019

Dosen Pembimbing,



Ainal Yaqin, M.Kom.
NIK. 190302255

PENGESAHAN

SKRIPSI

IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK KLASIFIKASI KOMENTAR SPAM PADA INSTAGRAM

yang dipersiapkan dan disusun oleh

Beta Priyoko

15.11.8912

telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Februari 2019

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Yuli Astuti, M.Kom.
NIK. 190302146



Dina Maulina, M.Kom.
NIK. 190302250



Ainul Yaqin, M.Kom.
NIK. 190302255

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 25 Februari 2019

DEKAN FAKULTAS ILMU KOMPUTER



Krisnawati, S.Si., M.T.
NIK. 190302038

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya pribadi.

Yogyakarta, 19 Februari 2019

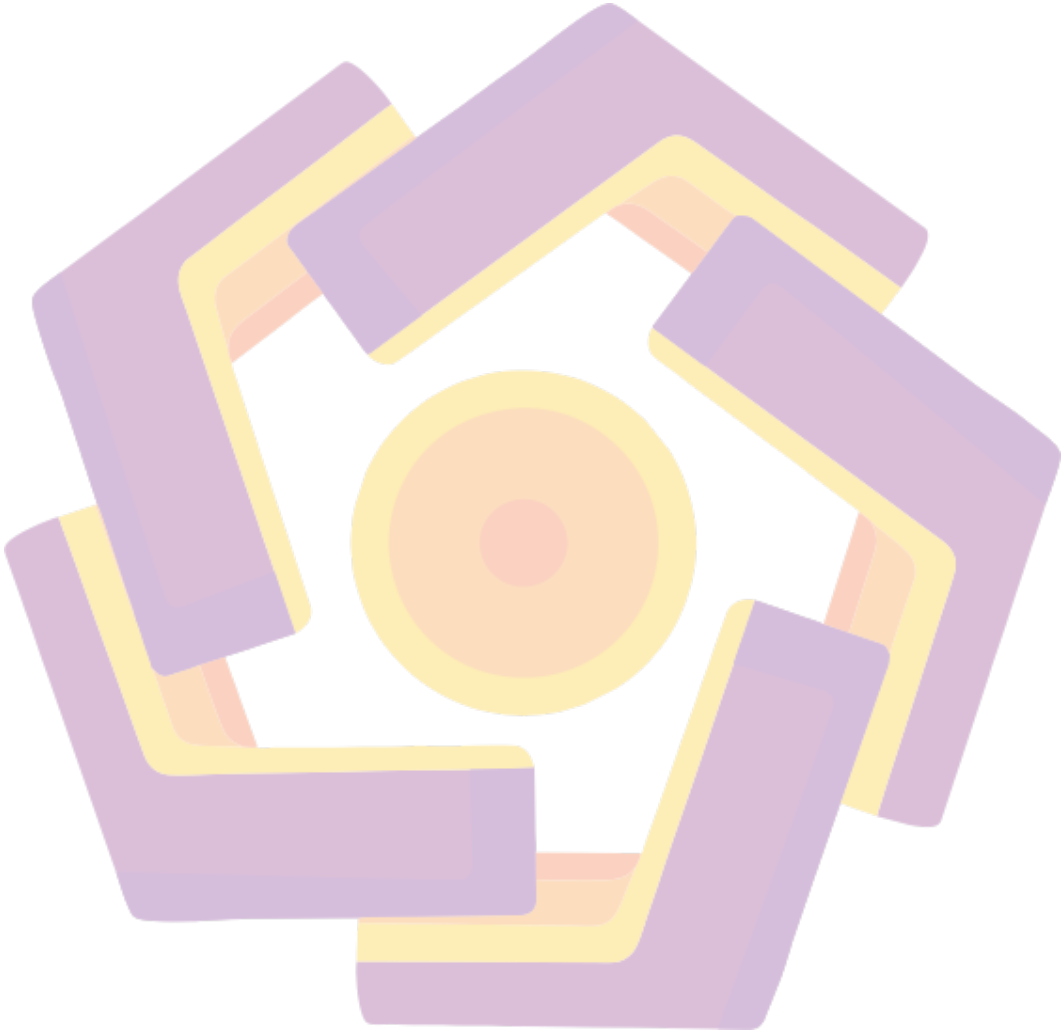


Beta Priyoko
NIM. 15.11.8912

MOTTO

“Pusing adalah bukti bahwa kita masih memiliki kepala.”

“Jangan seperti tempurung, didalamnya ada katak” – Cak Lontong



PERSEMBAHAN

Saya mempersembahkan skripsi ini kepada semua pihak yang terlibat secara langsung maupun tidak langsung dalam proses pembuatan skripsi.

1. Tuhan Yang Maha Esa yang memberikan segala nikmat dan kasih sayangnya sampai sejauh ini.
2. Kedua orang tua saya dan keluarga, yang selalu mendoakan, selalu menyemangati dan memberikan dukungan baik secara material maupun non material.
3. Bapak Ainul Yaqin, M.Kom. yang telah membimbing saya dari awal sampai akhir pembuatan skripsi.
4. Dosen-dosen Universitas AMIKOM Yogyakarta yang telah memberikan banyak ilmu selama kuliah.
5. Teman-teman diskusi dan sharing, Oky Antoro, Yulianto Mustaqim, dan Dicky Henry Saputra.
6. Teman skripsi saya, Muhammad Nurwahid, M Misbah Musthofa, Refianto Yusuf, Iqbal Tawakal Reno, Siti, dan Ola yang memberikan solusi, tempat bertanya ketika dimasa sulit dan tempat untuk ngeprint naskah hingga pembuatan skripsi selesai.
7. Teman-teman kelas IF06 2015 yang selalu menemani perkuliahan, mendukung dan memberikan semangat sampai saat ini. Semoga kita selalu bahagia dan menjadi pribadi yang lebih baik lagi.
8. Teman-teman dari softwareseni terutama grup cah javascript tapi tidak membahas javascript, Nico, Agita, Fandy dan Edim yang selalu mau sharing ilmunya kepada saya.

KATA PENGANTAR

Puji dan syukur saya panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat, hidayah dan kekuatan sehingga saya dapat menyelesaikan skripsi yang berjudul Implementasi Algoritma *Naive Bayes* untuk Klasifikasi Komentar Spam pada Instagram.

Skripsi ini saya buat guna menyelesaikan studi jenjang Strata Satu (S1) pada program studi Informatika fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta. Selain itu juga merupakan suatu bukti bahwa mahasiswa telah menyelesaikan kuliah jenjang program strata satu dan untuk memperoleh gelar Sarjana Komputer. Dengan selesainya skripsi ini, maka pada kesempatan ini saya mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Krisnawati, S.Si, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Bapak Ainul Yaqin, M.Kom. selaku dosen pembimbing yang selalu bijaksana memberikan bimbingan, nasehat serta waktunya selama penulisan skripsi ini.
4. Dosen Penguji dan segenap Dosen serta Karyawan Universitas AMIKOM Yogyakarta yang telah berbagi ilmu dan pengalamannya.
5. Kedua orang tua saya yang telah mendoakan, mendukung dan memberikan semangat kepada saya.
6. Orang-orang baik yang senantiasa mendoakan, memberi semangat dan bantuan kepada saya.
7. Teman-teman diskusi dan sharing, Oky Antoro, Yulianto Mustaqim, dan Dicky Henry Saputra.
8. Teman-teman Softwareseni yang selalu memberikan dukungan dan motivasi kepada saya.
9. Teman-teman IF06 angkatan 2015 yang telah menemani selama proses perkuliahan.

10. Semua pihak yang telah membantu baik secara langsung maupun tidak langsung.

Semoga Allah subhanahu wa ta'ala memberikan balasan yang lebih kepada semua yang telah ikut membantu saya hingga menyelesaikan skripsi ini. Demi perbaikan selanjutnya, saran dan kritik yang membangun diterima dengan senang hati dan rasa terima kasih. Semoga skripsi ini dapat bermanfaat bagi saya dan kita semua.

Yogyakarta, 19 Februari 2019

Beta Priyoko

DAFTAR ISI

HALAMAN JUDUL	I
PERSETUJUAN	II
PENGESAHAN	III
PERNYATAAN	IV
MOTTO	V
PERSEMBAHAN	VI
KATA PENGANTAR	VII
DAFTAR ISI	IX
DAFTAR TABEL	XI
DAFTAR GAMBAR	XII
INTISARI	XIV
<i>ABSTRACT</i>	XV
BAB I PENDAHULUAN	1
1.1 LATAR BELAKANG	1
1.2 RUMUSAN MASALAH	3
1.3 BATASAN MASALAH	3
1.4 MAKSUD DAN TUJUAN PENELITIAN	4
1.5 METODE PENELITIAN	4
1.6 SISTEMATIKA PENULISAN	6
BAB II LANDASAN TEORI	7
2.1 KAJIAN PUSTAKA	7
2.2 DASAR TEORI	13
2.2.1 <i>DATA MINING</i>	13
2.2.2 <i>METODE NAIVE BAYES</i>	14
2.2.3 <i>SPAM</i>	16

2.2.4	<i>K-FOLD CROSS VALIDATION</i>	17
2.2.5	<i>DATA FLOW DIAGRAM(DFD)</i>	17
2.2.7	<i>STEMMING</i>	19
2.2.8	<i>KLASIFIKASI</i>	19
2.2.9	<i>CONFUSION MATRIX</i>	19
BAB III ANALISIS DAN PERANCANGAN		22
3.1	PROFIL PERUSAHAAN INSTAGRAM	22
3.2	ANALISIS MASALAH.....	23
3.3	HASIL ANALISIS.....	23
3.4	DESKRIPSI SISTEM	24
3.5	PERHITUNGAN MANUAL <i>NAIVE BAYES</i>	26
3.5.1	<i>PROSES TRAINING</i>	26
3.5.2	<i>PROSES KLASIFIKASI</i>	30
3.6	ANALISIS KEBUTUHAN PERANGKAT LUNAK.....	32
3.6.1	<i>DEFINISI FITUR</i>	32
3.7	DFD (<i>DATA FLOW DIAGRAM</i>).....	33
3.8	RACANGAN ANTARMUKA PENGGUNA (<i>USER INTERFACE</i>).....	38
BAB IV IMPLEMENTASI DAN PEMBAHASAN		43
4.1	DATASET TRAINING	43
4.2	IMPLEMENTASI ANTARMUKA PENGGUNA	44
4.3	PEMBAHASAN <i>SOURCE CODE</i>	49
4.4	HASIL DAN ANALISIS PENGUJIAN MODEL	53
4.4.1	<i>PENGUJIAN MENGGUNAKAN 10-FOLD CROSS VALIDATION DAN CONFUSION MATRIX</i>	53
4.4.2	<i>HASIL PENGUJIAN</i>	55
BAB V PENUTUP		77
5.1	KESIMPULAN.....	77
5.2	SARAN.....	77
DAFTAR PUSTAKA		ERROR! BOOKMARK NOT DEFINED.

DAFTAR TABEL

Tabel 2.1 Perbedaan Penelitian.....	9
Tabel 2.2 <i>Data Flow Diagram(DFD)</i>	17
Tabel 2.3 <i>Confusion Matrix</i>	20
Tabel 3.1 <i>Data Training</i>	26
Tabel 3.2 Jumlah Dokumen.....	29
Tabel 3.3 Frekuensi Token.....	29
Tabel 3.4 Daftar Fitur Aplikasi.....	32
Tabel 4.1 Hasil <i>Confusion Matrix Fold 1</i>	56
Tabel 4.2 Hasil <i>Confusion Matrix Fold 2</i>	57
Tabel 4.3 Hasil <i>Confusion Matrix Fold 3</i>	59
Tabel 4.4 Hasil <i>Confusion Matrix Fold 4</i>	61
Tabel 4.5 Hasil <i>Confusion Matrix Fold 5</i>	63
Tabel 4.6 Hasil <i>Confusion Matrix Fold 6</i>	65
Tabel 4.7 Hasil <i>Confusion Matrix Fold 7</i>	67
Tabel 4.8 Hasil <i>Confusion Matrix Fold 8</i>	69
Tabel 4.9 Hasil <i>Confusion Matrix Fold 9</i>	71
Tabel 4.10 Hasil <i>Confusion Matrix Fold 10</i>	73
Tabel 4.11 Hasil Pengujian Model Menggunakan <i>10 Fold Cross Validation</i> dan <i>Confusion Matrix</i>	73
Tabel 4.12 Hasil Pengujian Menggunakan <i>Confusion Matrix</i> Secara Keseluruhan.....	75

DAFTAR GAMBAR

Gambar 3.1 Diagram Alir Klasifikasi.....	24
Gambar 3.2 Gambar Diagram Konteks	33
Gambar 3.3 DFD(<i>Data Flow Diagram</i>) Level 1	34
Gambar 3.4 DFD (<i>Data Flow Diagram</i>) Level 2 Proses Input Dataset	35
Gambar 3.5 DFD (<i>Data Flow Diagram</i>) Level 2 Proses <i>Training</i>	35
Gambar 3.6 DFD (<i>Data Flow Diagram</i>) Level 2 Proses Prediksi.....	35
Gambar 3.7 DFD (<i>Data Flow Diagram</i>) Level 2 Proses Validasi	36
Gambar 3.8 DFD (<i>Data Flow Diagram</i>) Level 3 Proses <i>Training</i>	36
Gambar 3.9 DFD(<i>Data Flow Diagram</i>) Level 3 Proses Prediksi.....	37
Gambar 3.10 DFD (<i>Data Flow Diagram</i>) Level 3 Proses Validasi	37
Gambar 3.11 Main Menu.....	38
Gambar 3.12 Halaman Prediksi	39
Gambar 3.13 Halaman Latih Model	40
Gambar 3.14 Halaman Validasi.....	41
Gambar 3.15 Halaman Hasil Validasi	42
Gambar 4.1 <i>Dataset Spam</i>	44
Gambar 4.2 <i>Dataset Notspam</i>	44
Gambar 4.3 Implementasi Main Menu.....	45
Gambar 4.4 Implementasi Halaman Prediksi	46
Gambar 4.5 Implementasi Halaman Latih Model.....	46
Gambar 4.6 Implementasi Halaman Validasi	47
Gambar 4.7 Implementasi Halaman Hasil Validasi.....	48
Gambar 4.8 <i>Source Code</i> Pengambilan Data Komentar Dari Instagram.....	49
Gambar 4.9 <i>Source Code</i> Latih Model	50
Gambar 4.10 <i>Source Code Text Preprocessing</i>	51
Gambar 4.11 <i>Source Code</i> Proses Klasifikasi	52
Gambar 4.12 Skenario Pengujian <i>10-Fold Cross Validation</i>	53
Gambar 4.13 <i>Pie Chart</i> Hasil Pengujian <i>Fold 1</i>	55
Gambar 4.14 <i>Pie Chart</i> Hasil Pengujian <i>Fold 2</i>	57
Gambar 4.15 <i>Pie Chart</i> Hasil Pengujian <i>Fold 3</i>	58

Gambar 4.16 *Pie Chart* Hasil Pengujian *Fold 4*60

Gambar 4.17 *Pie Chart* Hasil Pengujian *Fold 5*62

Gambar 4.18 *Pie Chart* Hasil Pengujian *Fold 6*64

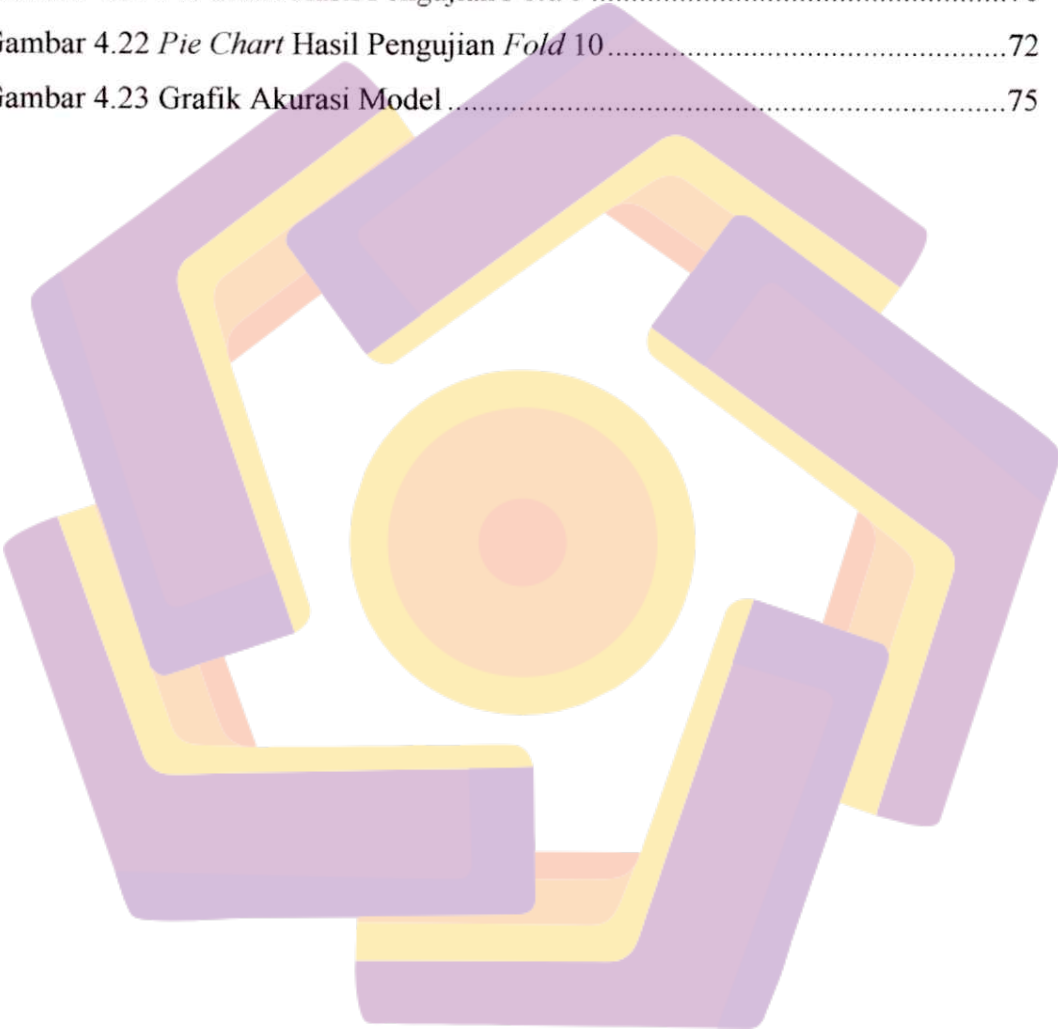
Gambar 4.19 *Pie Chart* Hasil Pengujian *Fold 7*66

Gambar 4.20 *Pie Chart* Hasil Pengujian *Fold 8*68

Gambar 4.21 *Pie Chart* Hasil Pengujian *Fold 9*70

Gambar 4.22 *Pie Chart* Hasil Pengujian *Fold 10*72

Gambar 4.23 Grafik Akurasi Model75



INTISARI

Instagram merupakan salah satu media sosial yang populer di Indonesia. Dengan Instagram, pengguna bisa membagikan momen kehidupannya dalam bentuk foto atau video. Para pengguna Instagram dapat saling mengikuti satu sama lain. Namun ketika seorang pengguna sudah mempunyai *followers* yang banyak, banyak juga akun Instagram yang menanggapi postingan dengan komentar yang dapat dikategorikan sebagai *spam*. Komentar *spam* biasanya banyak ditemui pada setiap postingan akun yang mempunyai *followers* banyak, terutama para *public figure* di Indonesia dan tentunya hal ini sangat mengganggu. Instagram telah menyediakan *services* untuk menghapus atau menyembunyikan komentar, akan tetapi masih diperlukan model untuk mendeteksi komentar itu *spam* atau *notspam*.

Algoritma *Naive Bayes* akan mencari probabilitas setiap kelas ketika komentar yang akan diprediksi diinputkan. Sebelum komentar di hitung nilai probabilitasnya pada masing-masing kelas, komentar akan diproses melalui tahap *preprocessing* yaitu *casefolding*, *cleaning*, *tokenizing*, dan *stemming*. Setelah diketahui nilai probabilitas dari masing-masing kelas, selanjutnya akan dibandingkan nilai probabilitasnya. Jika nilai probabilitas yang tertinggi adalah komentar yang dicek dengan hipotesa kelas *spam*, maka komentar itu dilabeli sebagai *spam*. Jika nilai probabilitas yang tertinggi adalah komentar yang dicek dengan hipotesa kelas *notspam*, maka komentar itu dilabeli sebagai *notspam*.

Kata kunci: *Naive Bayes, Bayesian, Text Classification, Preprocessing, Instagram*

ABSTRACT

Instagram is one of the most popular social media in Indonesia. With instagram, users can share their moments of life in the form of photos or videos. Instagram users can follow each other. But when a user already has many followers, many instagram accounts also respond to posts with comments that can be categorized as spam. Spam comments are usually found in every account post that has a lot of followers, especially public figures in Indonesia and of course this is very annoying. Instagram has provided services to delete or hide comments, but a model is still needed to detect comments that are spam or notspam.

The Naive Bayes algorithm will look for the probability of each class when the comments are inputted. Before the comments probability are calculated for each class, comments will be processed through the preprocessing stage, namely casefolding, cleaning, tokenizing, and stemming. After knowing the probability value of each class, then the probability value will be compared. If the highest probability value is a comment that is hypothesized as spam class, then the comment will be labeled as spam. If the highest probability value is a comment that is hypothesized as notspam class, then the comment will be labeled as notspam. this research has demonstrated the ability of Naive Bayes Algorithm to classify spam comments on instagram, obtained an accuracy of 80% with a balanced training dataset.

Keyword: *Naive Bayes, Bayesian, Text Classification, Preprocessing, Instagram*