

**PERBANDINGAN METODE NAIVE BAYES DAN METODE
CNN PADA KLASIFIKASI BERITA BBC NEWS**

SKRIPSI

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



disusun oleh

AKBAR NOOR HARTANTO

20.11.3484

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**PERBANDINGAN METODE NAIVE BAYES DAN METODE
CNN PADA KLASIFIKASI BERITA BBC NEWS**

SKRIPSI

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



disusun oleh

AKBAR NOOR HARTANTO

20.11.3484

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN METODE NAIVE BAYES DAN METODE CNN PADA
KLASIFIKASI BERITA BBC NEWS**

yang disusun dan diajukan oleh

Akbar Noor Hartanto

20.11.3484

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 23 agustus 2024

Dosen Pembimbing,



Rumini, M. Kom
NIK. 190302246

HALAMAN PENGESAHAN
SKRIPSI
PERBANDINGAN METODE NAIVE BAYES DAN METODE CNN PADA
KLASIFIKASI BERITA BBC NEWS

yang disusun dan diajukan oleh

Akbar Noor Hartanto

20.11.3484

Telah dipertahankan di depan Dewan Penguji
pada tanggal 23 agustus 2024

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Melany Mustika Dewi
NIK. 190302455

Ike Verawati, M.Kom
NIK. 190302237

Rumini, M.Kom
NIK. 190302246



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 23 agustus 2024

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Akbar Noor Hartanto
NIM : 20.11.3484

Menyatakan bahwa Skripsi dengan judul berikut:

PERBANDINGAN METODE NAIVE BAYES DAN METODE CNN PADA KLASIFIKASI BERITA BBC NEWS

Dosen Pembimbing : Rumini, 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, 23 agustus 2024

Yang Menyatakan,



Akbar Noor Hartanto

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur kepada Allah SWT, atas segala rahmat dan karunia-Nya, penulis dapat skripsi ini yang berjudul "Perbandingan Metode Naive Bayes dan Metode CNN pada Klasifikasi Berita BBC News". Perjalanan menyelesaikan skripsi ini tidaklah mudah, namun berkat dukungan dan doa dari orang-orang yang mendukung penulis, akhirnya skripsi ini dapat terselesaikan. Oleh karena itu, kupersembahkan dengan penuh rasa hormat dan terima kasih skripsi ini untuk:

1. Papah dan Mamah serta Eyang tercinta, Terima kasih atas doa, cinta, dukungan, dan pengorbanan yang tiada henti. Setiap langkahku adalah berkat doa kalian.
2. Kakak-kakakku tercinta, Yang selalu memberikan motivasi, semangat, dan dukungan yang berarti dalam setiap prosesku. Terima kasih atas kebersamaan dan cinta kasih kalian.
3. Kepada Ibu Rumini, M.Kom selaku Dosen Pembimbing, Yang telah dengan sabar memberikan arahan, bimbingan, dan dukungan akademik hingga terselesaikannya skripsi ini.
4. Seluruh Staf Dosen Yang telah memberikan ilmu dan pengalaman berharga selama masa studi, sehingga aku bisa mencapai titik ini.
5. Teman-teman kelas 20IF03, Kebersamaan dan dukungan kalian sangat berarti dalam perjalanan ini. Terima kasih atas kenangan, persahabatan, dan semangat yang terus menyala.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini yang berjudul "Perbandingan Metode Naive Bayes dan Metode CNN pada Klasifikasi Berita BBC News". Skripsi ini diajukan sebagai salah satu syarat untuk menyelesaikan studi Strata 1 di Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Dalam penyusunan skripsi ini, penulis menyadari bahwa tanpa bantuan, bimbingan, dan dukungan dari berbagai pihak, penulis tidak akan dapat menyelesaikan skripsi ini dengan baik. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof Dr. M. Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta
3. Ibu Windha Mega Pradnya D, M.Kom selaku ketua program studi SI Informatika.
4. Ibu Rumini, M.Kom selaku dosen pembimbing.
5. Bapak dan Ibu dosen Universitas Amikom Yogyakarta yang telah mengajar dan memberikan ilmu pengetahuan selama dibangku perkuliahan.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi perbaikan di masa yang akan datang. Semoga skripsi ini dapat memberikan manfaat bagi pembaca dan pengembangan ilmu pengetahuan, khususnya di bidang klasifikasi berita dengan metode Naive Bayes dan CNN.

Yogyakarta, September 2024

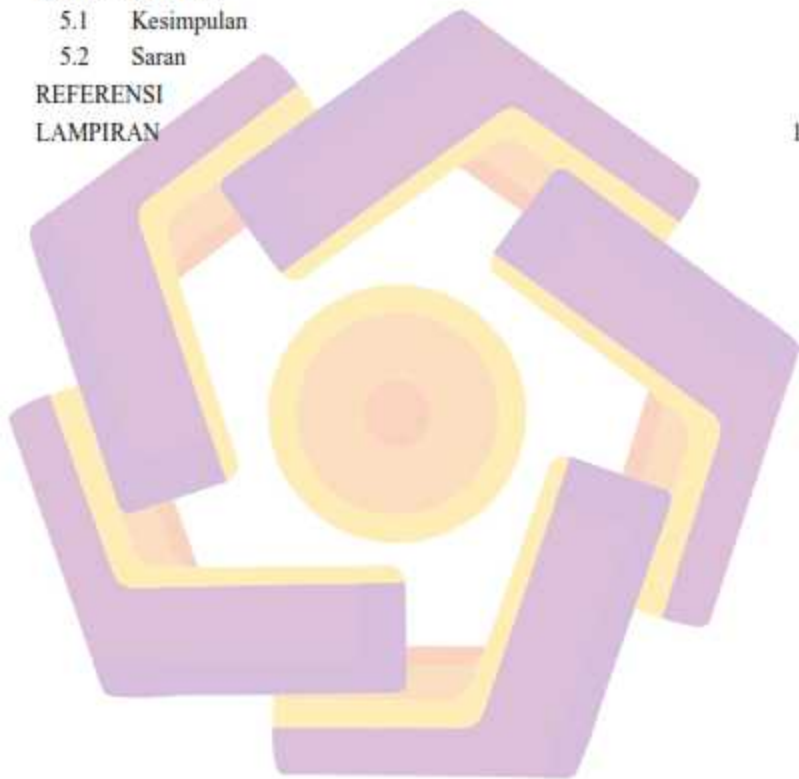
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI	xvi
<i>ABSTRACT</i>	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori	12
2.2.1 <i>Python</i>	12
2.2.2 <i>Machine Learning</i>	12
2.2.3 <i>Deep Learning</i>	12
2.2.4 <i>Text Mining</i>	13
2.2.5 <i>Stopwords</i>	13
2.2.6 <i>Data Training</i>	14
2.2.7 <i>Train-Test Split</i>	15

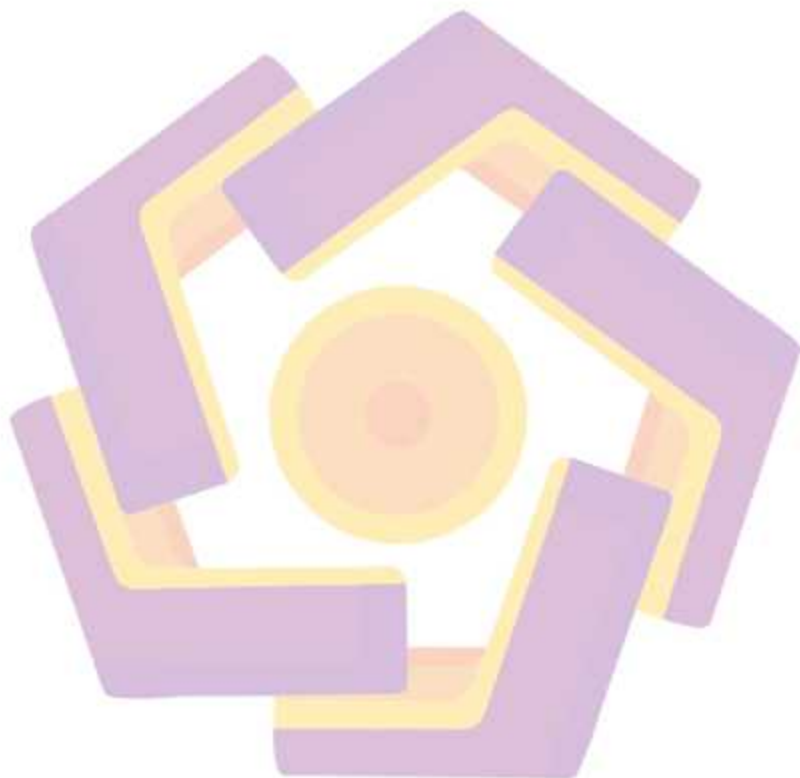
2.2.8.	<i>One-Hot Encoding</i>	15
2.2.9.	<i>Tokenize dan Encode</i>	15
2.2.10.	<i>Naïve Bayes</i>	16
2.2.11.	<i>Sigmoid</i>	19
2.2.12.	<i>Convolutional Neural Network (CNN)</i>	19
2.2.13.	<i>Confusion Matrix</i>	22
BAB III METODE PENELITIAN		25
3.1	Objek Penelitian	25
3.2	Alur Penelitian	26
3.2.1.	Pengumpulan data	27
3.2.2.	<i>Exploratory Data Analysis (EDA)</i>	27
3.2.3.	<i>Pre-processing Data</i>	29
3.2.3.1.	<i>Stopwords</i>	29
3.2.3.2.	<i>Data Training</i>	29
3.2.4.	Penerapan Model	32
3.2.4.1.	<i>Naïve Bayes</i>	32
3.2.4.2.	<i>Convolutional Neural Network (CNN)</i>	33
3.2.5.	Evaluasi Model	36
3.2.6.	Perbandingan	36
3.3	Alat dan Bahan	37
BAB IV HASIL DAN PEMBAHASAN		38
4.1.	Pengumpulan Data	38
4.2.	<i>Exploratory Data Analysis (EDA)</i>	39
4.3.	<i>Pre-processing data</i>	43
4.3.1.	<i>Stopwords</i>	43
4.3.2.	<i>Data Training</i>	47
4.3.2.1.	<i>Train-Test Split</i>	47
4.3.2.2.	<i>One-Hot Encoding</i>	51
4.3.2.3.	Normalisasi	61
4.3.2.4.	<i>Tokenize</i>	64
4.3.2.5.	<i>Encode</i>	68
4.3.2.6.	<i>Padding</i>	69
4.4.	Penerapan Model	70
4.4.1.	Naïve Bayes	70

4.4.2. Convolutional Neural Network (CNN)	81
4.5. Evaluasi Model	91
4.5.1. Naïve Bayes	91
4.5.2. Convolutional Neural Network (CNN)	93
4.6. Hasil Perbandingan Evaluasi	95
BAB V PENUTUP	97
5.1 Kesimpulan	97
5.2 Saran	97
REFERENSI	98
LAMPIRAN	101



DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian.....	9
Tabel 2.2 Confusion Matrix.....	23
Tabel 4.1 Tabel dataset	39

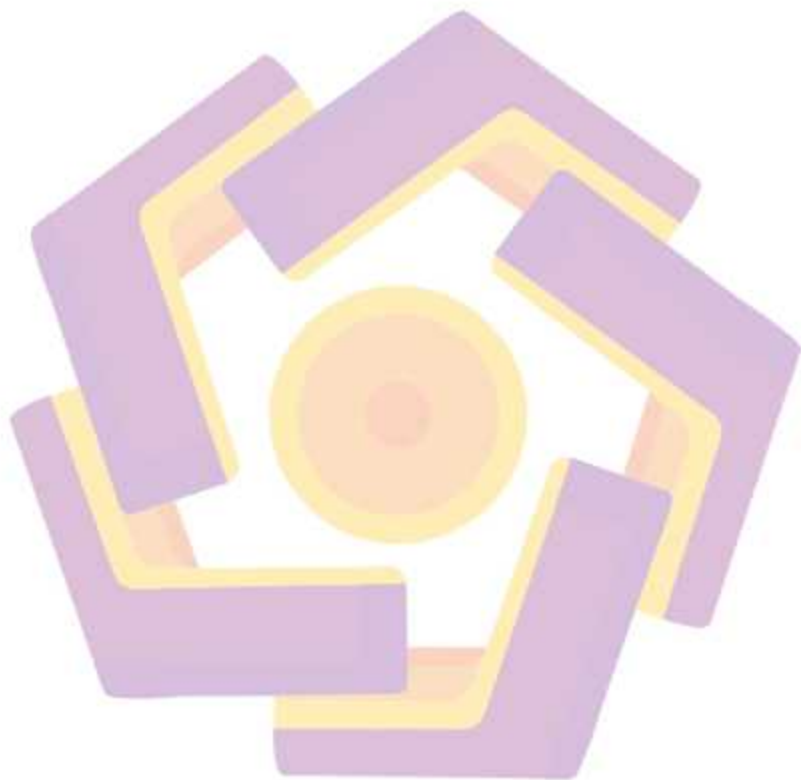


DAFTAR GAMBAR


Gambar 2.1 Ilustrasi CNN	20
Gambar 3.1 Alur penelitian.....	26
Gambar 3.2 Ilustrasi <i>Train-Test Split</i>	30
Gambar 4.1 <i>Data point</i> dan <i>Features</i>	40
Gambar 4.2 Kategori unik dalam kolom fitur <i>category</i>	40
Gambar 4.3 Hasil pembersihan baris dan kolom	41
Gambar 4.4 <i>Code</i> penggabungan kolom.....	42
Gambar 4.5 Hasil penggabungan kolom.....	42
Gambar 4.6 <i>Distribution of Datapoint Per Category</i>	43
Gambar 4.7 <i>Code</i> proses <i>stopword</i>	44
Gambar 4.8 Proses <i>stopword</i>	44
Gambar 4.9 Hasil proses <i>stopword</i>	47
Gambar 4.10 <i>Code Train-Test Split</i>	48
Gambar 4.11 <i>X Train</i> dan <i>Y Train</i>	48
Gambar 4.12 <i>X CV</i> dan <i>Y CV</i>	48
Gambar 4.13 Hasil <i>Train-Test Split</i>	49
Gambar 4.14 Chart Bar Data latih (Train data).....	50
Gambar 4.15 Chart Bar Data Validasi Silang (<i>cross-validation data/cv data</i>)... 50	
Gambar 4.16 <i>Code</i> proses <i>One-Hot Encoding</i>	51
Gambar 4.17 Hasil tokenisasi data latih	53
Gambar 4.18 Hasil tokenisasi data validasi silang	54
Gambar 4.19 Daftar <i>Vocabulary</i>	54
Gambar 4.20 <i>One-Hot Encode Matrix Train Data</i>	61
Gambar 4.21 <i>One-Hot Encode Matrix CV Data</i>	61
Gambar 4.22 <i>Code</i> proses normalisasi	61
Gambar 4.23 Normalisasi <i>data train</i>	64
Gambar 4.24 Normalisasi data validasi silang	64
Gambar 4.25 <i>Code</i> proses <i>Tokenize</i>	65
Gambar 4.26 Indeks <i>Vocabulary</i>	67
Gambar 4.27 <i>Sequence</i> hasil <i>Tokenize</i>	68
Gambar 4.28 <i>Code</i> proses <i>Encode</i>	68
Gambar 4.30 <i>Code</i> proses <i>Padding</i>	69
Gambar 4.31 Hasil proses <i>Padding</i>	70

Gambar 4.32 Code dan hasil Probabilitas Prior	70
Gambar 4.33 Code dan hasil Probabilitas Likelihood	72
Gambar 4.34 Code dan hasil Probabilitas Posterior	73
Gambar 4.35 Code dan Kalibrasi Sigmoid	75
Gambar 4.36 Code proses Logloss	76
Gambar 4.37 Daftar LogLoss	77
Gambar 4.38 Grafik Cross-Validation Error	78
Gambar 4.39 Grafik Cross-Validation Error	79
Gambar 4.40 Daftar hasil klasifikasi prediksi data silang dengan metode Naïve Bayes	80
Gambar 4.41 Gambar Akurasi model Naive Bayes	81
Gambar 4.42 Code proses Convolutional Neural Network (CNN)	81
Gambar 4.43 Code menampilkan hasil Embedding Layer	83
Gambar 4.44 Code menampilkan hasil Conv1D Layer	84
Gambar 4.45 Code menampilkan hasil GlobalMaxPooling1D Layer	85
Gambar 4.46 Code menampilkan hasil Embedding Layer	86
Gambar 4.47 Code menampilkan hasil Embedding Layer	86
Gambar 4.48 Code menampilkan hasil Embedding Layer	87
Gambar 4.49 Grafik loss model Convolutional Neural Network (CNN)	88
Gambar 4.50 Grafik Akurasi model Convolutional Neural Network (CNN)	89
Gambar 4.51 Daftar hasil klasifikasi prediksi kategori dengan metode Convolutional Neural Network (CNN)	90
Gambar 4.52 Gambar Akurasi model Convolutional Neural Network (CNN)	90
Gambar 4.53 Confusion Matrix Naïve Bayes	91
Gambar 4.54 Precision Matrix (Column Sum=1) Naïve Bayes	91
Gambar 4.54 Recall Matrix (Row Sum=1) Naïve Bayes	92
Gambar 4.55 Classification Report Naïve Bayes	92
Gambar 4.56 Confusion Matrix Convolutional Neural Network (CNN)	93
Gambar 4.57 Precision Matrix (Column Sum=1) Convolutional Neural Network (CNN)	94
Gambar 4.58 Recall Matrix (Row Sum=1) Naive Bayes	94
Gambar 4.59 Classification Report Convolutional Neural Network (CNN)	95

DAFTAR LAMPIRAN



DAFTAR LAMBANG DAN SINGKATAN



CNN	<i>Convolutional Neural Network</i>
NB	<i>Naïve Bayes</i>
BBC	<i>British Broadcasting Corporation</i>
NLP	<i>Natural Language Processing</i>
DM	<i>Data Mining</i>
ML	<i>Machine Learning</i>
DL	<i>Deep Learning</i>
LSTM	<i>Long Short-Term Memory</i>
RNN	<i>Recurrent Neural Network</i>
EDA	<i>Exploratory Data Analysis</i>
Σ	Total nilai
α	Nilai alpha
Π	Jumlah kali

DAFTAR ISTILAH

Precision	Besaran seberapa akurat prediksi
Recal	Besaran seberapa baik model
F1-Score	Rata-rata harmonik dari presisi dan recall
Tokenize	Memecah teks menjadi kata-kata
Token	Kumpulan kata kata
Encode	Proses mengubah token menjadi representasi numerik
Vektor	Representasi numerik hasil dari Encode
Index	Posisi dari token atau kata tertentu dalam vektor



INTISARI

Pada era digital ini, klasifikasi berita menjadi hal yang penting dalam mengelola dan menyajikan informasi yang melimpah, terutama dari sumber-sumber berita online seperti BBC News. Penelitian ini membandingkan dua metode klasifikasi teks, yaitu *Naïve Bayes* dan *Convolutional Neural Network (CNN)*, dalam mengklasifikasikan berita dari BBC News ke dalam lima kategori: bisnis, teknologi, politik, olahraga, dan hiburan. Dataset yang digunakan terdiri dari 2.225 berita yang telah diunduh dari *Kaggle*.

Hasil penelitian menunjukkan bahwa metode CNN memiliki keunggulan dalam hal akurasi, presisi, recall, dan F1-score dibandingkan dengan *Naïve Bayes*. Namun, *Naïve Bayes* tetap relevan untuk kasus-kasus tertentu yang membutuhkan metode yang lebih sederhana dan cepat. Penelitian ini memberikan wawasan tentang kelebihan dan kekurangan dari kedua metode klasifikasi dan diharapkan dapat menjadi referensi dalam pengembangan metode klasifikasi berita di masa depan.

Kata kunci: Klasifikasi Berita, *Naïve Bayes*, *Convolutional Neural Network*, BBC News, Machine Learning.

ABSTRACT

In this digital era, news classification has become essential in managing and presenting abundant information, particularly from online news sources like BBC News. This study compares two text classification methods, Naive Bayes and Convolutional Neural Network (CNN), in classifying BBC News articles into five categories: business, technology, politics, sports, and entertainment. The dataset used comprises 2,225 articles downloaded from Kaggle.

The results indicate that the CNN method outperforms Naive Bayes in terms of accuracy, precision, recall, and F1-score. However, Naive Bayes remains relevant for specific cases that require a simpler and faster method. This study provides insights into the strengths and weaknesses of both classification methods and is expected to serve as a reference for future news classification method development.

Keyword: *News Classification, Naïve Bayes, Convolutional Neural Network, BBC News, Machine Learning.*