

ANALISIS DAN PENERAPAN DATAMINING UNTUK MENDETEKSI BERITA PALSU (FAKE NEWS) PADA SOCIAL MEDIA DENGAN MEMANFAATKAN MODUL SCIKIT LEARN

Oleh :

Yosua Riadi Silitonga riadiyos@gmail.com

Pembimbing I : Ir. Munawar, MMSI, M.Com, Ph.D

Pembimbing II : Indriani Noor Hapsari, S.T, M.T

Abstrak - Saat ini media sosial tidak hanya sebagai media untuk berkomunikasi saja, namun sudah digunakan untuk kepentingan lainnya, seperti menerima, menyebarkan dan mengomentari informasi yang didapat. Facebook dan Twitter merupakan media sosial yang masih banyak digunakan saat ini. Dalam penggunaannya, kedua media sosial ini banyak digunakan untuk menyebarkan informasi yang dimana informasi yang disebarkan terkadang belum tentu benar adanya. Dalam penelitian ini dilakukan analisis apakah suatu informasi yang ada pada suatu media sosial bersifat benar atau palsu. Langkah pertama yang dilakukan dalam penelitian ini adalah pengumpulan data berupa berita atau opini menggunakan aplikasi Data Miner yang merupakan ekstensi Google Chrome dan Twitter Archiver merupakan pengaya Google Spreadsheet. Selanjutnya dilakukan pemodelan yang dimulai dengan *pre – processing* data dan terakhir menganalisis data. Data yang telah diproses menggunakan TF – IDF Vectorizer dan Count Vectorizer kemudian diproses kembali dengan menggunakan metode *scikit learn* menggunakan klasifikasi Multinomial Naive Bayes, Passive Agressive, dan Support Vector Classifier dengan menggunakan vector TF – IDF. Setelah proses klasifikasi dilakukan, hasil dari proses klasifikasi dan TF – IDF adalah kata yang sering muncul pada berita palsu dan berita nyata dari hasil klasifikasi. Dari hasil vektor tersebut dapat disimpulkan pemodelan ini dapat digunakan untuk menunjang pendeteksian berita palsu atau tidak dengan cara mengidentifikasi apakah kata hasil vektor dan klasifikasi terdapat pada berita yang telah teridentifikasi palsu pada dokumen berita yang telah dikumpulkan sebelumnya.

Kata kunci : *Media Sosial, Berita Palsu, Data Mining, Scikit - Learn*

Abstract - *Nowadays social media is not only a medium for communication, but also for other information, such as receiving, distributing and commenting on information obtained. Facebook and Twitter are social media that are still very popular. In its use, these two social media are widely used to integrate information in which the information disseminated is not necessarily true. In this study an analysis of whether information contained in a social media is true or false. The first step taken in this study is collecting data in the form of news or opinions using the Data Miner application which is an extension of Google Chrome and Twitter Archiver is an add-on to Google Spreadsheets. Furthermore, modeling is started with pre-processing data and finally analyzing data. Data that has been processed using TF - IDF Vectorizer and Count Vectorizer are then reprocessed using the scikit learn method using the Naive Bayes, Passive Agressive, and Support Vector Classifier Multinomial classifications using the TF - IDF vector. After the classification process is carried out, the results of the classification process and TF - IDF are words that often appear in fake news and real news from the classification results. From the results of these vectors it can be concluded that this modeling can be used to support the detection of false news or not by identifying whether the vector result word and classification are contained in the news that has been identified false on the news documents that have been collected previously.*

Kata kunci : *Social Media, Fake News, Data Mining, Scikit - Learn.*

Latar Belakang

Saat ini internet sudah menjadi kebutuhan yang harus dimiliki setiap kalangan masyarakat mulai dari kalangan tingkat atas serta kalangan tingkat menengah ke bawah dapat dengan mudah mengakses internet. Dengan kemudahan masyarakat mengakses internet, masyarakat harus mulai cerdas dalam menerima konten – konten yang diakses melalui internet. Hal ini dikarenakan mudahnya akses internet serta banyaknya konten – konten yang tidak dapat diuji kebenarannya yang dimana seluruh konten tersebut berisi tentang hal – hal yang dipenuhi dengan kebohongan yang berarti jenis penipuan dalam bentuk pernyataan yang tidak benar, terutama dengan maksud tujuan untuk menipu orang lain, atau biasa disebut dengan hoax serta ujaran kebencian.

Dalam penelitian yang dilakukan oleh Allcott Hunt, dkk (2017) mendefinisikan konten hoax atau pemberitaan palsu adalah berita yang sengaja dan dapat diverifikasi salah, yang bisa menyesatkan para pembaca. Pengertian lain mengenai pemberitaan palsu lainnya ialah artikel berita yang sengaja dan tidak benar palsu dan bisa menyesatkan pembaca (Shu Kai, dkk, 2017). Ada dua ciri utama definisi ini : keaslian dan niat. Pertama, berita palsu mencakup informasi palsu yang dapat diverifikasi seperti itu. Kedua, berita palsu dibuat dengan niat tidak jujur untuk menyesatkan konsumen. Definisi ini telah banyak diadopsi dalam penelitian terbaru. Definisi berita palsu yang lebih luas berfokus pada keaslian atau maksud konten berita.

Sosial Media merupakan sekelompok aplikasi berbasis internet yang membangun fondasi ideologi dan teknologi web 2.0 dan yang memungkinkan penciptaan dan pertukaran konten buatan pengguna . Web 2.0 mengacu pada platform di mana konten dan aplikasi terus diperbaharui oleh semua pengguna secara partisipatif dan kolaboratif (Laurell Christofer, 2014).

Twitter adalah layanan jejaring sosial dan mikroblog daring yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter, yang dikenal dengan sebutan kicauan (Tweet). Tweet adalah teks status pengguna yang digunakan untuk memberikan informasi maupun

opini. Pengguna Twitter, berdasarkan data PT Bakrie Telecom, memiliki 19,5 juta pengguna di Indonesia dari total 500 juta pengguna global (Kominfo, 2013). Kebanyakan pengguna Twitter di Indonesia adalah konsumen, yaitu yang tidak memiliki blog atau tidak pernah mengupload video di Youtube namun sering update status di Twitter dan Facebook.

Saat ini media - media yang bersifat tradisional dan modern yang berisi konten hoax sudah banyak tersebar melalui situs – situs, media cetak dan juga media sosial. Terlebih lagi jumlah aduan mengenai hoax dan SARA lebih besar dari pada pornografi yang dimana laporan tertinggi kedua ialah mengenai berita palsu. Tingkat laporan pada Januari 2017 mengenai SARA mencapai 5142, yang diikuti dengan berita palsu atau hoax sebanyak 5070 sementara disaat yang sama laporan seputar pornografi hanya sekitar 308 aduan (Damar Agustinus Mario, 2017).

Pengumpulan data dilakukan melalui teknik mining dengan aplikasi pihak ketiga yang dimana kita harus terhubung dengan API (Application Programming Interface) dari Twitter dan Facebook. API (Application Programming Interface) merupakan abstraksi yang menetapkan bagaimana satu program perangkat lunak dapat meminta layanan dari program perangkat lunak lain. Secara sederhana API dapat diartikan sebagai cara bagi dua perangkat lunak untuk berkomunikasi. API terdiri dari seperangkat aturan yang harus diikuti programmer untuk memanfaatkan program pengembang. API berfungsi untuk memfasilitasi interaksi antara kedua program (Wagner, Mathew C, dkk, 2013).

Setelah terhubung dengan API (Application Programming Interface) maka dapat dilakukan pengambilan data. Setelah pengambilan data dilakukan maka dapat dilakukan tahap penganalisa terhadap data yang telah diambil. Dalam penelitian yang telah dilakukan pada jurnal Fake News Detection on Social Media : A Data Mining Perspective, cara menganalisa berita palsu ialah dengan melakukan karakterisasi akun yang dianggap berbahaya. Dan cara mendeteksi ialah dengan menganalisa konten berita dan konteks sosial (Shu Kai, dkk, 2017).

Dengan memanfaatkan data data dari twitter dan facebook serta sumber berita dari website lainnya sebagai penunjang data, dapat diterapkan metode data mining untuk melihat perbedaan antara berita yang asli dengan berita yang dianggap palsu / tidak benar adanya. Berdasarkan latar belakang, maka akan dilakukan penelitian bagaimana teknik untuk membandingkan berita palsu dengan berita nyata yang mempunyai keterkaitan yang dimana data tersebut akan dilakukan perbandingan untuk dilakukan penarikan kesimpulan mengenai kebenaran suatu berita.

Identifikasi Masalah

Adapun identifikasi masalah yang akan dibahas pada penelitian “Analisis Sentimen Publik Terhadap Hasil Pemilihan Kepala Daerah Jawa Barat 2018 berdasarkan Opini Pengguna Twitter dengan Metode *Naive bayes classifier*“ ini adalah bagaimana melakukan teknik data mining untuk dapat mengetahui perbandingan suatu berita yang dianggap benar dengan berita yang dianggap palsu.

Batasan Masalah

Agar permasalahan yang tercakup dalam latar belakang tidak melebar dan menyimpang terlalu jauh dari tujuan, maka dibuat batasan masalah dari penelitian ini sebagai berikut :

1. Menampilkan informasi hasil proses dan analisis yang telah dilakukan terhadap data berita palsu dan berita nyata.
2. Mengumpulkan data berita palsu dan data berita nyata yang berasal dari sosial media.
3. Melakukan proses data yang telah dikumpulkan serta melakukan analisis dengan menggunakan teknik – teknik yang akan digunakan.
4. Membuktikan apakah teknik yang dipakai sudah efektif dalam mendeteksi berita bersifat palsu atau tidak

Tujuan Penelitian

Tujuan dari dilakukannya penelitian ini adalah :

1. Untuk mengetahui bagaimana kebenaran suatu berita dengan menggunakan teknik data mining.
2. Membandingkan dan menganalisis hasil klasifikasi berita palsu dengan berita lainnya yang berbeda.

LANDASAN TEORI

Definisi Berita

Berita merupakan cerita atau keterangan mengenai kejadian atau peristiwa yang hangat, laporan atau pemberitahuan atau pengumuman (Kbbi, 2017). Berita adalah akun aktivitas manusia yang tidak diterbitkan, yang berusaha untuk menarik, menginformasikan, atau mendidik para pembaca. Sebuah berita merupakan laporan dari suatu peristiwa dan kejadian itu sendiri yang dapat diartikan bahwa berita merupakan catatan peristiwa yang telah terjadi di era tertentu. (Farooq Umar, 2015)

Berita palsu atau yang dikenal dengan berita hoax juga dapat berarti artikel berita yang sengaja dan bisa dibilang salah, dan bisa menyesatkan pembaca. Mengacu pada literatur jurnalistik berdasarkan hukum, terdapat istilah berita hoax yang disebut dengan libel yaitu berita bohong yang berisikan tentang penghinaan, penistaan, pencemaran nama baik, hasutan, dan lain sebagainya yang merugikan orang lain yang dituangkan dalam tulisan dan slander yaitu secara lisan (Mardani Roland, 2017)

Definisi berita palsu yang lebih luas berfokus pada keaslian atau maksud konten berita. Dalam beberapa makalah menganggap berita satire sebagai berita palsu karena isinya palsu meskipun berita satire berorientasi pada hiburan dan mengungkapkan anggapannya tersendiri kepada konsumen (Shu Kai, dkk, 2017).

Menurut data science central masalah selanjutnya adalah berita palsu tidak semuanya sama. Oleh karena itu diperlukan model terpisah untuk mengidentifikasi target dalam kategori yang berbeda. Oleh karena itu Brenna Borlaug mengatasi masalah dengan mengklasifikasikan seperti ini (Vorhies William, 2017) :

1. Click Bait

Headline yang mengejutkan dimaksudkan untuk menghasilkan klik untuk

meningkatkan pendapatan iklan. Seringkali cerita – cerita ini sangat dibesar – besarkan atau salah sama sekali.

2. Propaganda

Artikel yang menyesatkan atau menipu dengan sengaja dimaksudkan untuk mempromosikan agenda penulis : Seringkali berisi penuh kebencian.

3. Komentar / Opini

Reaksi seseorang terhadap kejadian terkini. Artikel – artikel ini seringkali memberitahu pembaca bagaimana merasakan kejadian kejadian baru saat ini.

4. Humor / Satir

Artikel yang ditulis untuk hiburan. Cerita – cerita yang tidak dimaksudkan untuk dianggap serius

Dalam proses pendeteksian terbagi menjadi 2 bagian yaitu (Shu Kai, dkk ,2017) :

1. News Content (Konten Berita)

Knowledge-Based (Berbasis Pengetahuan)

Pendekatan berbasis pengetahuan mempunyai tujuan untuk menggunakan sumber eksternal untuk memeriksa fakta – fakta yang diajukan dalam konten berita. Tujuan pengecekan fakta untuk menetapkan nilai kebenaran pada klaim dalam konteks tertentu.

Style-based (Berbasis Gaya)

Pendekatan berbasis gaya mencoba mendeteksi berita palsu dengan menangkap manipulator dalam gaya penulisan konten berita.

2. Konteks Sosial

Dalam pendeteksian konteks sosial terbagi menjadi 2 bagian yaitu :

Berbasis Posisi

Dalam konteks ini memanfaatkan sudut pandang pengguna dari isi posting yang relevan untuk menyimpulkan kebenaran suatu artikel berita asli. Posisi posting pengguna dapat direpresentasikan secara eksplisit atau implisit. Posisi eksplisit adalah ekspresi langsung dari emosi seperti like atau dislike yang diungkapkan di Facebook. Sedangkan implisit dapat secara otomatis diekstrak dari posting media sosial. Deteksi posisi menentukan otomatis dari posting apakah pengguna mendukung, netral terhadap beberapa entitas target, acara, atau ide.

Berbasis Propagasi

Pendekatan berbasis propagasi berkaitan dengan posting media sosial yang relevan untuk memprediksi kredibilitas suatu berita. Asumsi dasarnya ialah bahwa kredibilitas suatu acara berita sangat terkait dengan kredibilitas posting media sosial yang relevan.

Narasi selama ini selalu dikaitkan dengan dongeng, cerita rakyat, atau cerita fiktif lainnya. Karena itu, analisis narasi selama ini banyak dipakai untuk mengkaji cerita fiksi. Padahal, narasi juga bisa dikaitkan dengan cerita yang berdasarkan fakta seperti berita. Dengan demikian analisis naratif juga bisa dipakai untuk menganalisis teks berita yang diangkat dari suatu fakta.

Berita juga merupakan suatu narasi. Ini berarti berita mengikuti atau memenuhi syarat - syarat sebagai suatu narasi. Yang pertama, Rangkaian peristiwa pada berita umumnya terdiri atas sejumlah peristiwa yang dirangkai menjadi suatu berita. Berita hampir tidak mungkin hanya mengangkat satu peristiwa. Yang kedua, Rangkaian peristiwa berita pada dasarnya juga mengikuti jalan cerita atau logika tertentu. Yang ketiga, pada dasarnya berita juga bukan copy paste dari realitas. Realitas yang luas dan kompleks tidak mungkin diberitakan sama persis. Dalam konteks seperti ini ada peristiwa yang dimasukkan, dan ada peristiwa yang diuang karena tidak sesuai dengan jalan cerita yang hendak disampaikan oleh jurnalis. Analisis naratif adalah analisis mengenai narasi, baik narasi fiksi (novel, puisi, cerita rakyat, dongeng, komik dll) ataupun fakta – seperti berita (Eriyanto,2013).

Sosial Media

Menurut jurnal John Fotis, Media sosial didefinisikan sebagai kelompok aplikasi berbasis internet yang membangun fondasi ideologi dan teknologi dan yang memungkinkan pembuatan dan pertukaran konten yang dibuat oleh pengguna (Fotis John, dkk , 2017).

Menurut Hiroshi Ishikawa dalam bukunya Social Big Data Mining sosial media dapat diklasifikasikan menjadi beberapa

kategori berdasarkan layanan yang disediakan (Ishikawa Hiroshi, 2015) :

1. Blogging : Layanan pada kategori ini memungkinkan pengguna untuk mempublikasikan penjelasan, sentimen, evaluasi, tindakan, dan gagasan mengenai topik tertentu termasuk acara personal maupun sosial kedalam bentuk teks dengan gaya buku harian,
2. Micro Blogging : Pengguna menjelaskan topik tertentu sering dalam bentuk teks pendek pada micro blogging. Contohnya ialah, Tweet yang merupakan artikel dari Twitter yang terdiri paling banyak 140 karakter.
3. SNS (Social Network Service) : Layanan pada kategori ini secara harfiah mendukung jaringan sosial antara pengguna.
4. Sharing Service : Layanan pada kategori ini memungkinkan pengguna membagikan video, audio, foto dan bookmarks.
5. Social News : Melalui layanan ini pengguna dapat memberikan berita sebagai sumber utama dan juga dapat memposting ulang dan mengevaluasi item berita favorit setelah diposting.

Facebook merupakan sebuah layanan jejaring sosial yang diluncurkan pada bulan Februari 2004 hingga 2012. Facebook memiliki lebih dari satu miliar pengguna aktif, lebih dari separuhnya menggunakan telepon genggam. Facebook merupakan situs jejaring sosial gratis yang populer yang memungkinkan pengguna terdaftar membuat profil, mengunggah foto dan video, mengirim pesan dan tetap berhubungan dengan teman, keluarga dan rekan kerja. Situs yang tersedia dalam 37 bahasa berbeda, termasuk fitur publik seperti itu (Rouse Margaret, 2017).

Pada saat ini Mining Data melalui Facebook telah populer dan sangat berguna dalam beberapa tahun terakhir. Data yang digali akan bernilai dan konstruktif bagi komersial, ilmiah, dan banyak bidang prediksi dan analisis lainnya, terutama apabila data ini proses lebih dalam, seperti pembersihan data dan pembelajaran mesin. Facebook juga menyediakan sebuah situs penyediaan website yang mengizinkan bagi para developer untuk mengakses data mereka, sejak para pengantusias data mengungkapkan minta

yang begitu kuat terhadap data yang berasal dari Facebook. Website ini menyediakan banyak metode yang simpel dan mudah dipahami dengan petunjuk yang terperinci bagi

Merupakan layanan jejaring sosial dan mikroblog daring yang memungkinkan pengguna untuk mengirim dan membaca pesa berbasis teks hingga 140 karakter, yang dikenal dengan sebutan kicauan (tweet) (Twitter, 2013). Twitter dalam bahasa Inggris berarti berkicau, hal ini sesuai dengan fungsi situs Twitter yang memungkinkan anda untuk mengatakan apa saja yang sedang anda lakukan atau pikirkan kapan saja dan diketahui banyak orang. Jejaring sosial berlogo burung berwarna biru ini sebenarnya sama saja dengan jejaring sosial lainnya. Yang berbeda ialah jejaring sosial yang didirikan pada Maret 2006 oleh perusahaan rintisan Obvius Corp ini adalah sebuah layanan microblogging dimana anggotanya bisa menjawab pertanyaan, "Apa sih yang sedang Anda lakukan?", dengan mengirimkan pesan singkat terbatas hanya sampai 140 karakter. Twitter merupakan salah satu platform micro blogging yang didirikan oleh Jack Doursey pada tahun 2005 (Ishikawa Hiroshi, 2015).

Keunggulan Twitter terletak pada karakteristik format jawaban pendek yang disebut dengan tweet (Fakhrurroja Hanif, dkk, 2009). Twitter berorientasi pada teks seperti platform blogging pada umumnya seperti WordPress dan Blogger. Tentunya, tweet juga dapat menambahkan tautan untuk menuju ke media lain (Ishikawa Hiroshi, 2015).

Application Program Interface (API)

Application Program Interface (API) atau Antarmuka Pemrograman Aplikasi adalah abstraksi yang menentukan bagaimana satu program perangkat lunak dapat meminta layanan dari program perangkat lunak lain. Dalam istilah sederhana, API adalah cara untuk dua perangkat lunak untuk berkomunikasi. API terdiri dari seperangkat aturan yang harus diikuti oleh programmer untuk memanfaatkan program pengembang. Tujuan API adalah untuk memfasilitasi interaksi antara kedua program ini. API digunakan oleh semua jenis perangkat lunak. Contohnya layanan web seperti Twitter menggunakan API untuk mengizinkan aplikasi

pihak ketiga digunakan dan berinteraksi dengan data mereka. API dimaksudkan agar mudah beradaptasi sehingga bisa digunakan dengan berbagai cara oleh pemrogram. Tanpa API, Twitter akan ditutup secara efektif dari aplikasi pengembang luar (Wagner Matthew C, dkk, 2013).

Data Mining

Data mining diartikan sebagai proses menemukan pola dalam data. Proses harus otomatis atau semiotomatis. Pola – pola yang ditemukan harus menghasilkan beberapa keuntungan. Contohnya ialah dalam keuntungan ekonomi (Chakrabarti Soumen, 2009).

Data mining merupakan suatu istilah yang diciptakan untuk menggambarkan proses peralihan melalui database yang besar untuk mencari pola yang menarik dan yang sebelumnya tidak diketahui. Dengan aksesibilitas dan berlimpahnya data saat ini, membuat data mining menjadi suatu hal yang sangat penting dan diperlukan. atau juga dapat diartikan sebagai suatu proses logikal yang digunakan untuk mencari dari sejumlah data untuk mendapatkan data yang berguna (Ramageri Bharati M, 2010). Singkatnya data mining adalah untuk menemukan pola dan struktur bermakna yang muncul dalam sejumlah besar data yang digunakan oleh aplikasi (Ishikawa Hiroshi, 2015).

Data mining dapat dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan yaitu :

1. Deskripsi
Peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.
2. Estimasi
Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori.
3. Prediksi
Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang.
4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori.

5. Pengklusteran

Metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik antara satu data dengan data yang lain. Clustering merupakan salah satu metode data mining yang bersifat tanpa arahan.

6. Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam suatu waktu.

Data mining dapat diposisikan dalam konteks teknologi yang lebih luas disebut dengan KDD (Knowledge Discovery in Databases) atau dapat diartikan penemuan pengetahuan dalam basis data. Pada umumnya KDD mengambil data dari Datawarehouse atau Database dan menciptakan pengetahuan sebagai hasilnya. KDD terdiri dari beberapa langkah – langkah berikut :

1. Pembersihan data (Data Cleansing) : Menghapus data dari sumber data yang mengganggu konsistensi.
2. Integrasi Data (Data Integration) : Menyatukan dua atau lebih sumber data jika diperlukan dan disimpan dalam basis data
3. Pengurangan dan Pemilihan data : Pilih bagian – bagian penting dari data sebagai target penambangan data dari database. Mengurangi data target dengan jumlah data yang dapat diproses secara praktis.
4. Konversi Data : Mengubah data target menjadi struktur data yang sesuai untuk penambangan data.
5. Data Mining : Mengekstrak pola dari data yang dikonversi menggunakan metode.
6. Evaluasi pola : Evaluasi pola yang diekstraksi berdasarkan ukuran tertentu dan identifikasikan sebagai pengetahuan pola.
7. Representasi pengetahuan : Mengekspresikan dan memvisualisasikan yang teridentifikasi pengetahuan untuk presentasi yang efektif.

Oleh karena itu, Data Mining merupakan salah satu langkah penting dalam proses penemuan pengetahuan.

Scikit - Learn

Scikit – Learn adalah modul python yang mengintegrasikan berbagai algoritma pembelajaran mesin state-of-the-art untuk masalah yang diawasi dan tidak diawasi skala menengah. Paket ini berfokus pada membawa pembelajaran mesin ke non-spesialis menggunakan bahasa tingkat tinggi tujuan umum. Penekanan diberikan pada kemudahan penggunaan, kinerja, dokumentasi, dan konsistensi API. Ini memiliki ketergantungan minimal dan didistribusikan dibawah lisensi BSD yang disederhanakan, mendorong penggunaannya baik dalam aturan akademis dan komersial. (Pedregosa Fabiann, 2011). Library dibangun diatas SciPy (Scientific Python) yang harus diinstal sebelum menggunakan scikit – learn. Tumpukan ini meliputi :

1. Numpy: Paket array n – dimensi dasar
2. Scipy : Pustaka dasar untuk komputasi ilmiah
3. Matplotlib : Komprehensif 2D / 3D
4. Ipyton : Peningkatan konsol interaktif
5. Sympy : Matematika simbolik
6. Pandas : Struktur dan analisis data.

Ekstensi atau modul untuk Scipy care secara konvensional diberinama Scikits. Modul ini menyediakan algoritma pembelajaran dan diberi nama scikit – learn (Brownlee Jason, 2014).

METODOLOGI PENELITIAN

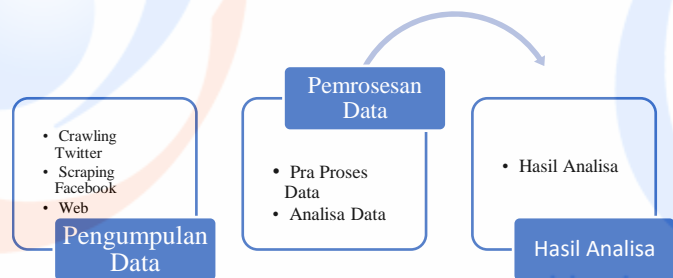
Metode Penelitian

Dalam penelitian ini, terdapat cara yang dilakukan dalam pengumpulan data terkait landasan teori, kepastian informasi, dan analisis masalah untuk memecahkan permasalahan dalam penelitian yaitu studi pustaka dan studi literatur.

1. Studi Literatur

Pada tahap ini dilakukan pengambilan data dengan cara membandingkan literatur sejenis, baik dari buku maupun dari lapangan yaitu berupa jurnal penelitian yang melakukan penelitian serupa sebelumnya. Data yang dihasilkan dari tahap ini berupa kelebihan dan kekurangan metode yang digunakan dari

penelitian sebelumnya dan metode mana yang cocok dengan penelitian terkait.



Gambar 1 Metode Penelitian

2. Penelitian

Penelitian ini memiliki tahapan yaitu sebagai berikut :

1. Tahap Pengumpulan Data

Pada tahap ini, dilakukan pengumpulan data yang diperoleh dari media sosial Twitter dan Facebook dengan menggunakan ekstensi pada Spread Sheet dan Google Chrome yaitu Twitter Archiver dan Data Mining yang ada pada aplikasi yang akan digunakan untuk menarik data. Pada proses pengumpulan data training menggunakan teknik mencari berita pada fan page atau grup pada media sosial Facebook yang memverifikasi berita palsu. Contoh fan page dan grup yang memverifikasi berita palsu seperti Masyarakat Anti Fitnah Indonesia, Turn Back Hoax dan Forum Anti Fitnah Hasut dan Hoax. Teknik yang digunakan oleh grup atau fan page tersebut untuk memverifikasi berita palsu adalah Knowledge – Based yaitu menggunakan sumber eksternal untuk memeriksa fakta – fakta yang diajukan dalam konten berita dan menggunakan teknik propagasi yaitu pendekatan dengan posting yang relevan pada media sosial untuk memprediksi kredibilitas suatu berita.

2. Tahap Pra - Proses Data

Pada tahap ini, data yang sudah dikumpulkan dari media sosial Twitter dan Facebook akan diklasifikasikan berdasarkan 2 bagian. Bagian pertama ialah data training yang dimana data tersebut berisi data berita yang telah

dinyatakan bersifat palsu. Bagian kedua ialah data testing yang dimana berisi data berita yang dimana belum diketahui apakah bersifat palsu atau tidak. Setelah data telah dikumpulkan dan diberi label, maka selanjutnya memasuki tahap pra proses yaitu dengan menggunakan teknik klasifikasi vektor yaitu CountVectorizer dan TF - IDF Vectorizer.

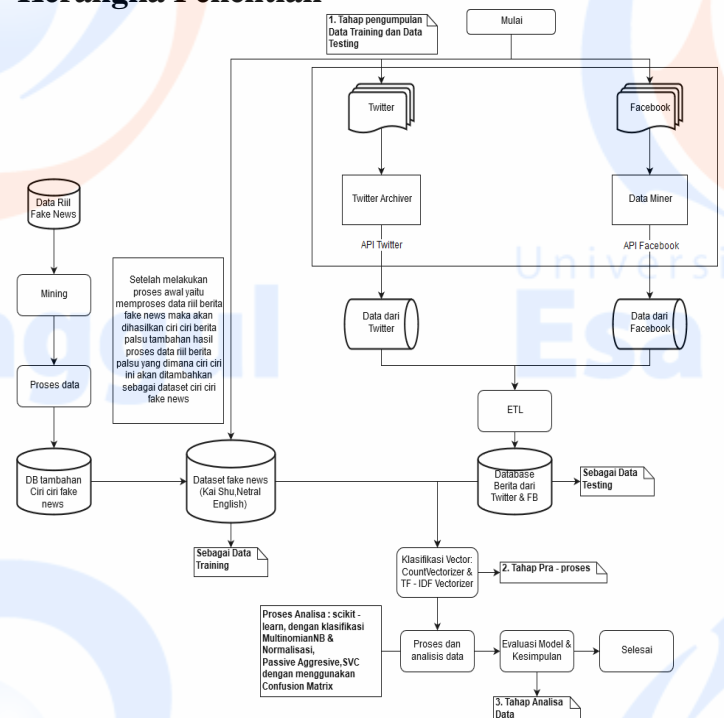
3. Tahap Analisis Data
Setelah melakukan klasifikasi vektor pada tahapan pra – proses, maka selanjutnya adalah dengan menguji permodelan test dengan menggunakan MultinomialNB. Pada tahapan ini akan menggunakan permodelan yang ada pada artikel datacamp (Jarmul Cathrine, 2017) yang berjudul Detecting Fake News with Scikit – Learn. Perbedaan yang akan dilakukan pada penelitian ini adalah dengan memunculkan normalisasi pada proses MultinomialNB dan menambahkan klasifikasi Support Vector Classifier.

Dalam tahapan ini akan membandingkan antara TF – IDF dengan Count Vectorizer. Untuk dapat melihat perbandingan lebih jelas maka dibuat Confusion Matriks yang ada pada dokumentasi scikit – learn dengan menggunakan klasifikasi MultinomialNBClassifier, Passive Agresive Classifier, dan Support Vector Classifier.

4. Tahap Penarikan Kesimpulan
Setelah data training dan data testing melewati tahap proses, dan telah memasuki tahap pengecekan permodelan dengan menggunakan confusion matriks dan klasifikasi yang digunakan, maka pada tahap ini akan dilakukan evaluasi model. Pada evaluasi model ini akan dilakukan proses secara bertahap yang dimana bertujuan untuk mengetahui apakah hasil dari permodelan mempunyai nilai yang berbeda setiap prosesnya dengan menggunakan jumlah data secara bertahap. Setelah melakukan evaluasi model maka selanjutnya akan dilakukan penarikan kesimpulan atas

hasil analisis yang telah dilakukan terhadap model.

Kerangka Penelitian



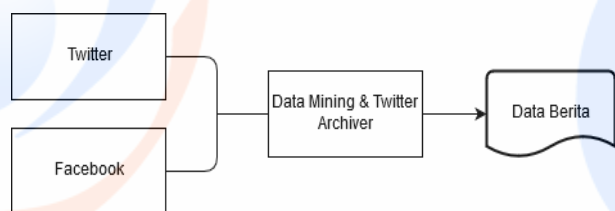
Gambar 2 Kerangka Penelitian

PERANCANGAN PEMODELAN

Untuk memulai pemodelan pada tahap pertama yang dilakukan adalah menginstall perangkat lunak Jupyter Notebook untuk mendukung metode scikit-learn dan terakhir merupakan keluarannya yang berupa confusion matrix yang dimana terlebih dahulu harus menginstall python versi 3.x dan dilengkapi dengan package lain seperti Scikit Learn, Numpy, Pandas, dan Matplotlib.

1. Pengumpulan Data

Pada tahap ini, dilakukan pengumpulan data yang diperoleh dari media sosial Twitter dan Facebook dengan menggunakan ekstensi pada Spread Sheet dan Google Chrome yaitu Twitter Archiver dan Data Mining yang ada pada aplikasi yang akan digunakan untuk menarik data.



Gambar 3 Proses pengambilan data

Hasil dari proses pengambilan data berupa file *csv* yang dimana akan digunakan dalam proses pemodelan.

2. Hasil Praproses Data

Sebelum memasuki tahap permodelan klasifikasi dan juga untuk mendapatkan kata apa yang paling banyak bersifat palsu dan benar, maka dilakukan pra - proses data terlebih dahulu. Pra - proses data yang dilakukan adalah dengan membangun pengklasifikasian vektor yang dimana proses klasifikasi vektor yang akan dilakukan menggunakan klasifikasi CountVectorizer dan TF - IDF Vectorizer. Praproses data ini dilakukan dengan menggunakan aplikasi Jupyter Notebook, dengan hasil praproses pada Tabel 3 berikut :

```
count_df.head()
```

	00	00web	05	061	07	075	09	090718	0945web	10	...	yogyakarta	yori	yudhoyono	yesri	yesrihza_mhd	zat	zionis	ojustru	cekainnya	celatar		
0	0	0	0	0	0	0	1	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0	
1	1	0	0	0	0	0	1	0	0	0	...	0	0	0	1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0	0

5 rows x 2304 columns

Gambar 3 Hasil Count Vectorizer

```
tfidf_df.head()
```

	00	00web	05	061	07	075	09	090718	0945web	10	...	yogyakarta	yori	yudhoyono	yesri	yesrihza_mhd	zat	zionis	ojustru				
0	0.000000	0.0	0.0	0.0	0.0	0.0	0.059214	0.0	0.0	0.0	...	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.046558	0.0	0.0	0.0	0.0	0.0	0.041670	0.0	0.0	0.0	...	0.0	0.0	0.046558	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

5 rows x 2300 columns

Gambar 4 Hasil TF - IDF Vectorizer

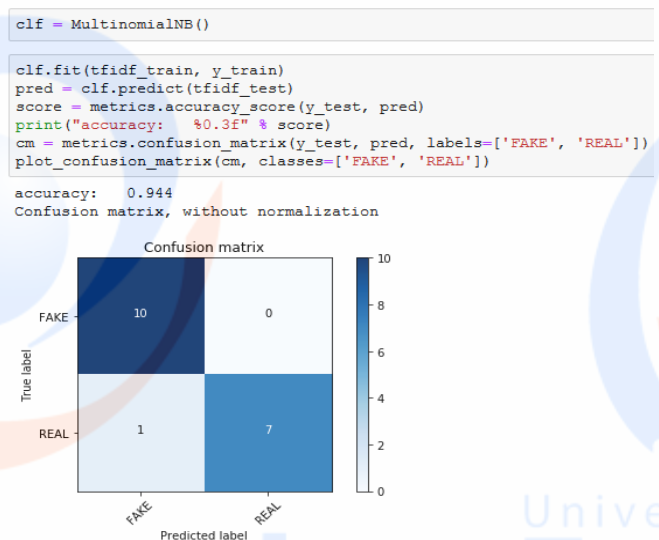
Pada Gambar 3 menunjukkan hasil dari CountVectorizer yang dimana hasil nilai dari perhitungan tersebut adalah jumlah kemunculan kata dari berita yang telah diproses. Pada Gambar 4 menunjukkan hasil dari TF - IDF yang dimana hasil dari nilai tersebut merupakan nilai pembobotan kata dari berita yang telah diproses.

3. Hasil Analisa Data

Setelah melakukan pra - proses data maka selanjutnya adalah analisis data. Untuk dapat menganalisa data, maka dilakukan proses klasifikasi dan membuat Confusion Matrix. Pada proses klasifikasi terdapat 3 klasifikasi yang digunakan dengan memakai dataset TF - IDF yaitu :

1. MultinomialNB (Multinomial Naive Bayes) dengan menggunakan CountVectorizer dan Normalisasi
2. PassiveAgressive Classifier
3. Support Vectori Classifier

Dikarenakan setiap akurasi mempunyai proses perhitungan yang berbeda - beda, maka tujuan digunakannya klasifikasi tersebut adalah mengetahui tingkat akurasi dari masing - masing klasifikasi dengan menggabungkan perhitungan klasifikasi dan TF - IDF yang dimana hasil tingkat akurasi dari setiap klasifikasi tersebut mempengaruhi tingkat keakuratan pada hasil vektor tertinggi. Hasil dari berbagai klasifikasi terdapat pada gambar berikut :



Gambar 5 Hasil MultinomialNB (TF -IDF Vector)

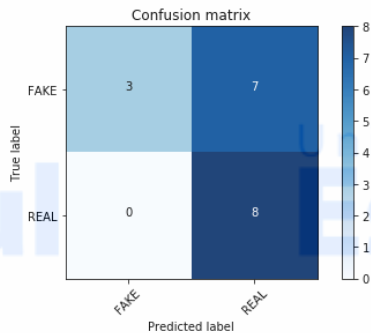
```

clf = MultinomialNB()

clf.fit(count_train, y_train)
pred = clf.predict(count_test)
score = metrics.accuracy_score(y_test, pred)
print("accuracy: %0.3f" % score)
cm = metrics.confusion_matrix(y_test, pred, labels=['FAKE', 'REAL'])
plot_confusion_matrix(cm, classes=['FAKE', 'REAL'])

```

accuracy: 0.611
Confusion matrix, without normalization



Gambar 6 Hasil MultinomialNB(Count Vector)

Pada Gambar 5 dan Gambar 6 merupakan hasil dari klasifikasi MultinomialNB yang menggunakan vektor TF – IDF dan Count Vectorizer. Hasil dari klasifikasi ini digunakan untuk mengetahui perbedaan tingkat akurasi antara TF - IDF dan CountVectorizer dengan menggunakan klasifikasi MultinomialNB.

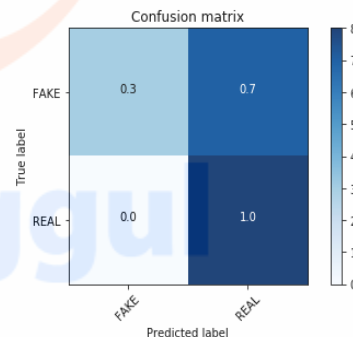
```

clf = MultinomialNB()

clf.fit(count_train, y_train)
pred = clf.predict(count_test)
score = metrics.accuracy_score(y_test, pred)
print("accuracy: %0.3f" % score)
cm = metrics.confusion_matrix(y_test, pred, labels=['FAKE', 'REAL'])
plot_confusion_matrix(cm, classes=['FAKE', 'REAL'], normalize=True)

```

accuracy: 0.611
Normalized confusion matrix



Gambar 8 Hasil MultinomialNB (Count Vector) Normalisasi

Pada Gambar 7 dan Gambar 8 merupakan hasil dari klasifikasi MultinomialNB yang menggunakan vektor TF – IDF dan Count Vectorizer dengan normalisasi. Fungsi dari normalisasi ini untuk mengetahui persentasi dari setiap kolom matrix.

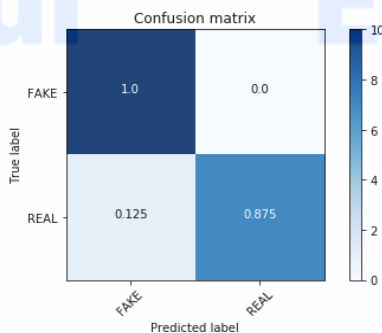
```

clf = MultinomialNB()

clf.fit(tfidf_train, y_train)
pred = clf.predict(tfidf_test)
score = metrics.accuracy_score(y_test, pred)
print("accuracy: %0.3f" % score)
cm = metrics.confusion_matrix(y_test, pred, labels=['FAKE', 'REAL'])
plot_confusion_matrix(cm, classes=['FAKE', 'REAL'], normalize=True)

```

accuracy: 0.944
Normalized confusion matrix



Gambar 7 Hasil MultinomialNB (TF – IDF Vector) Normalisasi

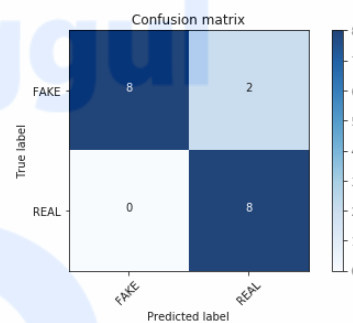
```

linear_clf = PassiveAggressiveClassifier(max_iter=50)

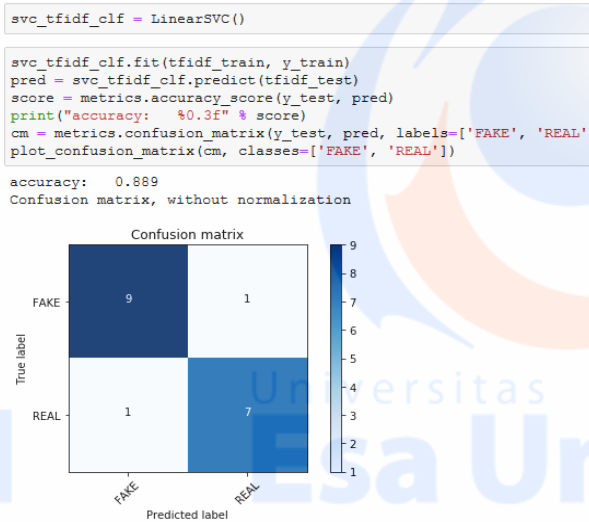
linear_clf.fit(tfidf_train, y_train)
pred = linear_clf.predict(tfidf_test)
score = metrics.accuracy_score(y_test, pred)
print("accuracy: %0.3f" % score)
cm = metrics.confusion_matrix(y_test, pred, labels=['FAKE', 'REAL'])
plot_confusion_matrix(cm, classes=['FAKE', 'REAL'])

```

accuracy: 0.889
Confusion matrix, without normalization



Gambar 9 Hasil Klasifikasi Passive Aggressive



Gambar 10 Hasil Klasifikasi SVC

Pada Gambar 9 dan Gambar 10 merupakan hasil dari klasifikasi Passive Agressive dan Support Vector Classifier yang menggunakan vektor TF – IDF. Klasifikasi ini digunakan karena mempunyai fungsi yang sama dalam mengolah data yang masif.

4. Pengujian Model

Setelah semua tahap klasifikasi selesai dilakukan, selanjutnya adalah menggunakan pengklasifikasian Passive Agressive Classifier dan dataset TF – IDF untuk memeriksa vektor teratas untuk berita palsu dan nyata. Hasil dari vektor tersebut disajikan pada gambar berikut :

```

FAKE -0.721190732274431 partai
FAKE -0.5661989235541037 kurma
FAKE -0.561118985003837 uang
FAKE -0.5316950396422145 berita
FAKE -0.5211271944748995 joko
FAKE -0.518676940358091 israel
FAKE -0.5130941681876131 dalam
FAKE -0.49713997691714357 foto
FAKE -0.49449058130352036 edy
FAKE -0.46087661310368 pare
FAKE -0.4262435534437425 satelit
FAKE -0.4070624994190463 negara
FAKE -0.39712873217228417 widodo
FAKE -0.37791438987439335 bahwa
FAKE -0.3742531370041207 miliar
FAKE -0.3729631092632747 roket
FAKE -0.3677424444766963 demokrat
FAKE -0.3628441829016275 ma
FAKE -0.36284183637672013 lapangan
FAKE -0.3438282740228293 ktp

```

Gambar 11 Hasil Vektor Fake teratas

```

REAL 0.724509688154381 ayam
REAL 0.5632814711647842 nu
REAL 0.5399022061875809 lift
REAL 0.5312313675949676 yg
REAL 0.4978654181015422 hari
REAL 0.4918161642174495 mandiri
REAL 0.4757330959682802 jadi
REAL 0.4418720885777058 pancasila
REAL 0.40816036978498854 biar
REAL 0.401772202224947 ke
REAL 0.3931203129635479 gak
REAL 0.39265614995460424 tombol
REAL 0.39186445240502754 warning
REAL 0.38608607604061784 presiden
REAL 0.3855132855463888 tau
REAL 0.37401301216251204 bumi
REAL 0.36914312693368545 gratis
REAL 0.36914312693368545 gb
REAL 0.36327179748969257 velodrome
REAL 0.36327179748969257 terbaik

```

Gambar 12 Hasil Vektor Real teratas

Pada Gambar 11 dan Gambar 12 menunjukkan hasil dari klasifikasi Passive Agressive Classifier dan TF – IDF yang berupa kata – kata dari hasil proses tersebut. Nilai dari setiap vektor berasal dari proses klasifikasi yang diurutkan dari hasil perhitungan tertinggi. Dari nilai tersebut dapat menunjukkan bahwa kata yang muncul pada vektor tersebut merupakan kata yang paling sering muncul dari dokumen yang telah diproses

EVALUASI MODEL

Pada tahapan Evaluasi model akan menggunakan data berita yang berbeda dengan data berita pada permodelan. Pada tahapan ini akan menguji evaluasi model untuk mengetahui apakah terdapat hal – hal yang dapat berpengaruh terhadap hasil akhir dari pemodelan yang telah dilakukan. Pada tahapan awal dilakukan tahap evaluasi dengan menggunakan 20 data. Hasil evaluasi model disajikan dalam tabel berikut :

Tabel 1 Hasil Evaluasi Model

Proses	Jumlah Data	Hasil Proses
1	20 Data Berita Data training : 10 data berita Data testing : 10 data berita	MultinomialNB (TF - IDF) : 57% MultinomialNB (CountVectorizer) : 100% PassiveAggressive

		Classifier : 71% Support Vector Classifier : 57%
2	30 Data Berita Data training : 15 data berita Data testing : 15 data berita	MultinomialNB (TF - IDF) : 30% MultinomialNB (CountVectorizer) : 60% PassiveAggressive Classifier : 40% Support Vector Classifier : 30%
3	38 Data Berita Data training : 20 data berita Data testing : 18 data berita	MultinomialNB (TF - IDF) : 92% MultinomialNB (CountVectorizer) : 54% PassiveAggressive Classifier : 85% Support Vector Classifier : 92%
	Jumlah Rata – rata	MultinomialNB (TF - IDF) : 60% MultinomialNB (CountVectorizer) : 71% PassiveAggressive Classifier : 65,3% Support Vector Classifier : 60%

Setelah data berita palsu dan data berita yang belum teridentifikasi palsu terkumpul dan diproses menggunakan klasifikasi vektor TF – IDF dan Count Vectorizer serta klasifikasi MultinomialNB, Passive Aggressive dan Support Vector Classifier, maka hasil yang didapat adalah tingkatan kata yang sering muncul pada berita yang telah diproses dari hasil klasifikasi. Dalam hal ini, dapat disimpulkan bahwa pemodelan ini dapat digunakan untuk menunjang pendeteksian berita palsu atau tidak dengan cara mengidentifikasi apakah kata hasil vektor dan klasifikasi terdapat pada berita yang telah teridentifikasi palsu pada dokumen berita yang telah dikumpulkan sebelumnya. Kata hasil vektor yang digunakan untuk mengidentifikasi tentunya harus menggunakan klasifikasi yang mempunyai tingkat akurasi yang tinggi agar hasil kata yang muncul pada hasil proses akan semakin baik.

Saran

Saran untuk penelitian selanjutnya :

1. Perlu menggunakan macam – macam klasifikasi agar dapat menghasilkan analisa yang berbeda dari setiap klasifikasi yang digunakan.
2. Pengembangan pemodelan yang sebelumnya agar dapat menghasilkan hasil yang lebih signifikan dengan cara melakukan komparasi dari setiap klasifikasi yang digunakan.
3. Penelitian selanjutnya bisa dikembangkan kembali dengan data yang sudah akurat terbukti palsu dan menggunakan jumlah data yang lebih banyak lagi dengan mencoba mengambil data dari media sosial lainnya agar mendapatkan hasil yang lebih signifikan.
4. Menambahkan variabel lainnya seperti tingkatan jumlah berita fake dan real yang berbeda pada evaluasi model untuk mendapatkan hasil yang lebih maksimal dan

Kesimpulan Evaluasi Model

Berdasarkan evaluasi model yang telah dilakukan terhadap klasifikasi yang telah dilakukan pada tahap pertama MultinomialNB 20 data mempunyai tingkat akurasi sebesar 57%, pada tahapan kedua dengan jumlah data 30 data mendapatkan tingkat akurasi sebesar 30% dan pada tahap ketiga dengan 38 data mendapatkan tingkat akurasi sebesar 92%. Dari tingkat akurasi tersebut dapat disimpulkan dengan menambahkan data berita belum tentu dapat meningkatkan tingkat akurasi pada proses pemodelan. Hal ini juga dapat berpengaruh dari isi berita pada dokumen yang diproses. Untuk mendapatkan tingkat akurasi yang tinggi maka diperlukan data berita palsu yang lebih banyak dari berita nyata.

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan pada proses awal pencarian data palsu teknik yang digunakan saat memilih data training yang berisi berita palsu adalah dengan menggunakan teknik *knowledge – based* dan teknik propagasi.

dapat mengevaluasi model lebih dalam.

DAFTAR PUSTAKA

- Allcott, Hunt, Matthew Gentzkow. 2017. Social Media and Fake News in the 2016 Election. *Jurnal : Journal of Economic Perspective – Volume 31, Number 2 – Spring 2017 – Pages 211 – 236*
- Akhyar, Turkhan. 2017. The Characteristics of Hoax News, According to a Legal Communications Expert. Diakses dari <http://www.en.netralnews.com/news/currentnews/read/5092/the.characteristics.of.hoax.news.according.to.a.legal.communications.expert> pada Oktober 2017
- Asmaya, Fela. 2015. Pengaruh Penggunaan Media Sosial Facebook Terhadap Perilaku Prosocial Remaja Di Kenagarian Koto Bangun. *Universitas Riau Jurnal Jom FISIP Volume 2 No.2*
- Chakrabarti, Soumen. 2009. *Data Mining : Know It All*. Morgan Kaufmann. USA
- Damar, Agustinus Mario. 2017. Jumlah Aduan Hoax dan SARA Lampau Konten Pornografi. Diakses dari <https://www.liputan6.com/teknoread/3053599/jumlah-aduan-hoax-dan-sara-lampau-konten-pornografi> pada Oktober 2017
- Defiyanti, Sofi, Mohamad Jajuli. 2015. Integrasi Metode Klasifikasi dan Clustering dalam Data Mining. *Jurnal Konferensi Nasional (KNIF) 2015*
- Eriyanto. 2013. Analisis Naratif: Dasar-dasar dan Penerapannya dalam Analisis Teks Berita Media Edisi Pertama. Kencana. Jakarta
- Fakhrurroja, Hanif, Aris Munandar. 2009. *Twitter Ngoceh Dapat Duit*. Great Publisher. Yogyakarta
- Farooq, Umar. 2015. What is News Definition and Source of News. Diakses dari <http://www.studylecturenotes.com/journalism-mass-communication/what-is-news-meaning-definition-and-sources-of-news>
- Fotis, John, dkk. 2012. Social media use and impact during the holiday travel planning process. *Jurnal Han, Jiawei, dkk. 2012. Data Mining : Concepts and techniques*. Morgan Kaufmann. USA
- Henning, Michi. 2007. *API Design Matters*. ACM Queue
- Hutahaean, Jeperson. 2014. *Konsep Sistem Informasi*. Deepulish. Yogyakarta
- Laurell, Christofer. 2014. *Commercialising Social Media – A Study Of Fashion (BlogoSpheres)*. US-AB. Stockholm
- Ishikawa, Hiroshi. 2015. *Social Big Data Mining*. Taylor & Francis Group. USA
- Kominfo. 2013. *Kominfo : Pengguna Internet di Indonesia 63 juta Orang*. Diakses dari https://kominfo.go.id/index.php/content/detail/3415/Kominfo+%3A+Pengguna+Internet+di+Indonesia+63+Juta+Orang/0/berita_satker pada Oktober 2017
- Mardani, Roland. 2017. Jenis macam berita jurnalistik, Diakses dari <http://www.bitebrands.co/2017/07/jenis-macam-berita-jurnalistik.html> pada Oktober 2017
- Obar, Jonathan A, Steve Wildman. 2015. *Social Media Definition and the Governance Challenge : An Introduction to the Special Issue*. Michigan State University
- Octoparse Team. 2017. *Facebook Data Mining*. Diakses dari <https://www.octoparse.com/blog/facebook-data-mining/> pada Oktober 2017

O'Keeffe, Gwenn Schurgin, dkk. Clinical Report-The Impact of Social Media on Children. Adolescents,an Families. 2011

Oracle. 2012. Oracle Database 2 Day + Data Warehousing Huide, 11g Release 2(11.2). Oracle America,Inc. USA

Ramageri, Bharati M. 2010. Data Mining Techniques and Applications. Jurnal : Indian Journal of Computer Science and Engineering Vol.1 No. 4 301 - 305

Rouse, Margaret. 2014. Facebook Definition. Diakses dari <https://whatis.techtarget.com/definition/Facebook> pada Oktober 2017

Sartika, Dewi, dkk. 2017. Perbandingan Algoritma Klasifikasi Naive Bayes,Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian.Universitas Indo Global Mandiri. Jurnal : Jatisi,Vol.1 No.2 Maret 2017

Shu, Kai. 2017. Fake News Detection on Social Media : A Data Mining Perspective. Michigan State University.USA

Singhal, Hanoop. 2007. Data Warehousing and Data mining Techniques fo Cyber Security. Springer. USA

Vorhies, William. 2017. Using Algorithms to Detect Fake News – The State of the Art. Diakses dari : https://www.datasciencecentral.com/profiles/blogs/using-algorithms-to-detect-fake-news-the-state-of-the-art?xg_source=activity pada November 2017

Wagner, Matthew C, dkk. 2013. Application Programming Interfaces (“APIs”) A Primer and Discussion of Oracle America v. Google. Jurnal