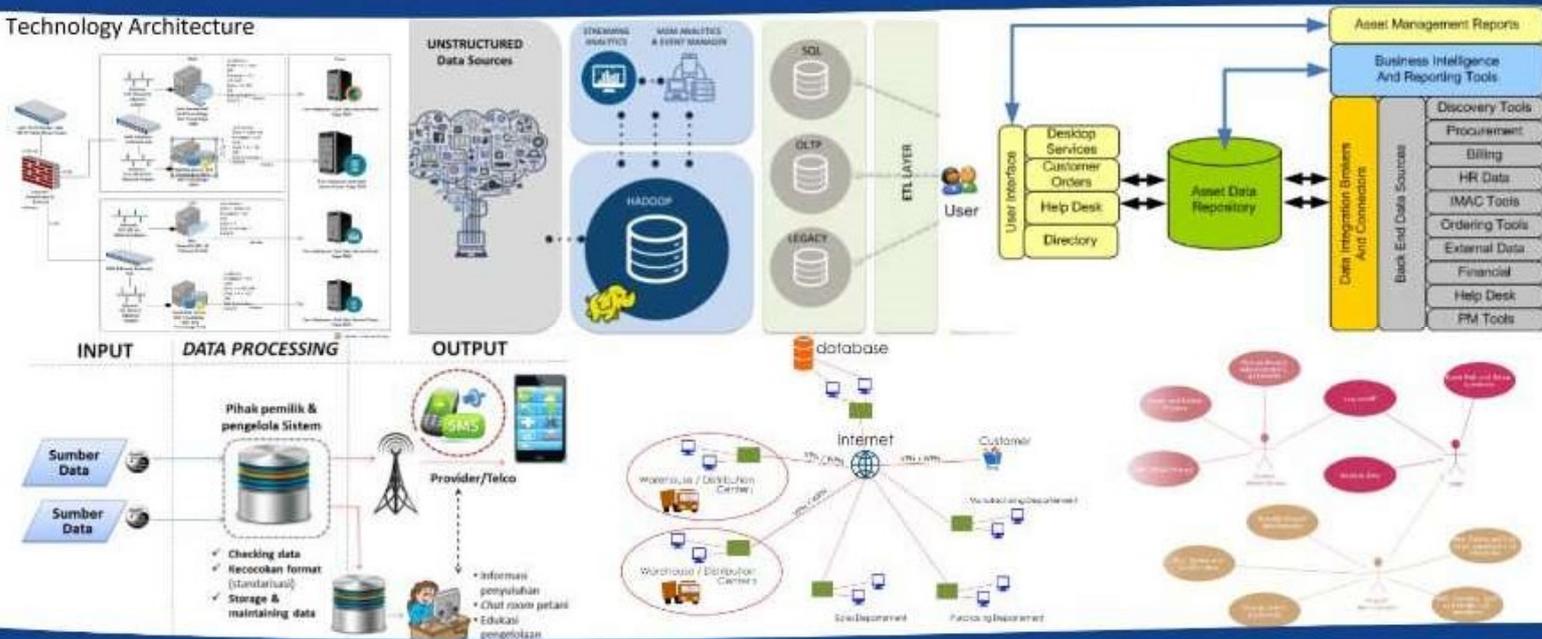


# MISI

## JURNAL MANAJEMEN INFORMATIKA & SISTEM INFORMASI



### Technology Architecture



Diterbitkan Oleh LPPM STMIK Lombok  
Jln. Basuki Rahmat No.105 Praya, Lombok Tengah - NTB  
Telp dan Fax (0370) 654310 - e-journal.stmiklombok.ac.id/jsi  
email. lppm@stmiklombok.ac.id





**DEWAN REDAKSI**

**JURNAL MISI ( JURNAL MANAJEMEN INFORMATIKA DAN SISTEM INFORMASI)**

**Jurnal Manager**

**Wire Bagye, S.Kom.,M.Kom** (STMIK Lombok, SINTA ID : 5992010)

**Reviewer :**

**Resad Setyadi, S.T., S.Si., MMSI., Ph.D** (cand)- Institut Teknologi Telkom Purwokerto

SCOPUS ID 57204172534, SINTA ID : 6113570

**Yesaya Tommy Paulus, S.Kom., MT., Ph.D.** - STMIK Dipanegara Makassar

SCOPUS ID 57202829909, SINTA ID : 6002004

**Lalu Mutawalli, S.Kom., M.I.Kom., M.Kom** - STMIK Lombok

SCOPUS ID : 57205057118, SINTA ID : 6659709

**Saruni Dwiasnati, ST., MM., M.Kom** - Universitas Mercu Buana

SCOPUS ID : 57210968603, SINTA ID : 6150854

**Ida Bagus Ary Indra Iswara, S.Kom., M.Kom** - STMIK STIKOM Indonesia

SCOPUS ID 57203711945, SINTA ID : 183498

**Erlin Windia Ambarsari** - Universitas Indraprasta PGRI

SCOPUS ID : 56242503900, SINTA ID : 5998887

**Wafiah Murniati, ST., MT.** - STMIK Lombok

SCOPUS ID : 56242503900, SINTA ID : 5998887

**Yuliadi, S.Kom., M.Kom** - Universitas Teknologi Sumbawa

SINTA ID : 6730786

**Fachrudin Pakaja, S.Kom, M.T** - Universitas Gajayana

SINTA ID : 6164357

**Ahmad Jufri, S.Kom., M.T** - Sekolah Tinggi Teknologi STIKMA Internasional

SINTA ID : 172241

**Mohammad Taufan Asri Zaen, ST., MT** - STMIK Lombok

SINTA ID : 5992087

**Hairul Fahmi, S.Kom., M.Kom** - STMIK Lombok

SINTA ID : 5983160

**I Ketut Putu Suniantara, S.Si., M.Si** - ITB STIKOM Bali

SINTA ID : 6086221

**Nawassyarif S. Kom., M.Pd.** - Universitas Teknologi Sumbawa

SINTA ID : 6722660

**Muhamad Malik Mutoffar, ST., MM., CNSS** - Sekolah Tinggi Teknologi Bandung

SINTA ID : 6013819

**Editor :**

**Saikin, Skom., M.Kom.** - STMIK Lombok

**Vrestanti Novalia Santosa, M.Pd.** - IKIP Budi Utomo Malang

**Desain Grafis & Web Maintenance**

**Jihadul Akbar, S.Kom** - STMIK Lombok

**Secretariat**

**Maulana Ashari, M.Kom** - STMIK Lombok



## DAFTAR ISI

<b>1</b>	HIGH AVAILABILITY DYNAMIC SHARDING DATABASE SERVER DENGAN METODE FAIL OVER DAN CLUSTERING <b>Afirda Desember Riawati<sup>1</sup>, M Irfan<sup>2</sup>, Khaeruddin<sup>3</sup>, Amrul Faruq<sup>4</sup></b>	<b>1 - 10</b>
<b>2</b>	IMPLEMENTASI METODE EXPONENTIAL SMOOTHING PADA SISTEM INFORMASI PERAMALAN STOK DI PT ATLANTIC BIRURAYA JOMBANG <b>Teguh Priyo Utomo<sup>1</sup>, Beda Puspita Chandra<sup>2</sup>, Febri Afriyan Pratama<sup>3</sup>, Ivan Dwi Fibrian<sup>4</sup></b>	<b>11 - 19</b>
<b>3</b>	ANALISIS SENTIMEN ULASAN EKSPEDISI J&T EXPRESS MENGGUNAKAN ALGORITMA NAIVE BAYES <b>Mahardika Tania Nitami<sup>1</sup>, Herny Februariyanti<sup>2</sup></b>	<b>20 - 29</b>
<b>4</b>	RANGKING INDEKS BERITA LARANGAN MUDIK DENGAN METODE TF-IDF DAN <i>COSINE SIMILARITY</i> MENGGUNAKAN <i>MACHINE LEARNING</i> <b>Muhammad Syahrani<sup>1</sup>, Kusnadi<sup>2</sup>, Bambang Joko Triwibowo<sup>3</sup>, Yusuf Arif Setiawan<sup>4</sup>, Fariszal Nova Arviantino<sup>5</sup>, Didi Rosiyadi<sup>6</sup></b>	<b>30 - 38</b>
<b>5</b>	PENGEMBANGAN APLIKASI PENILAIAN PEMBELAJARAN DARING (E-LEARNING) BERBASIS WEB <b>Muh Khatami Akib<sup>1</sup>, Ratna Shofiati<sup>2</sup>, Ahmad Zuhdi<sup>3</sup></b>	<b>39 - 47</b>
<b>6</b>	PENERAPAN <i>RESEARCH AND DEVELOPMENT</i> (R&D) DALAM MEMBANGUN ALAT PENYIRAMAN TANAMAN OTOMATIS BERBASIS ARDUINO <b>Khairul Imtihan<sup>1</sup>, Ernawati<sup>2</sup>, Lalu Mutawalli<sup>3</sup></b>	<b>48 - 55</b>
<b>7</b>	SISTEM LAYANAN LABORATORIUM BERBASIS WEBSITE LABORATORIUM JURUSAN SEJARAH UNNES <b>Junaidi Fery Lusianto<sup>1</sup>, Tsabit Azinar Ahmad<sup>2</sup>, Sulton Widiantoro<sup>3</sup>, Nawanggi Dwindi Arsila<sup>4</sup></b>	<b>56 - 64</b>
<b>8</b>	PREDIKSI PROSES PERSALINAN MENGGUNAKAN ALGORITMA KNN BERBOBOT PADA MONITORING ELEKTRONIK PERSONAL HEALTH RECORD IBU HAMIL <b>Sutrino<sup>1</sup>, Dwiati Wismarini<sup>2</sup></b>	<b>65 - 76</b>
<b>9</b>	ANALISIS SENTIMEN PERGURUAN TINGGI TERMEWAH DI INDONESIA MENURUT ULASAN GOOGLE MAPS MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) <b>Adhitia Erfina<sup>1</sup>, Neng Resti Wardani<sup>2</sup></b>	<b>77 - 85</b>
<b>10</b>	RANCANG BANGUN SISTEM COMPUTER BASED TEST UNTUK UJIAN SEMESTER MAHASISWA (STUDI KASUS : POLITEKNIK HASNUR) <b>Abdullah Ardi<sup>1</sup>, Achmad Rayhan Alief<sup>2</sup></b>	<b>86 - 94</b>
<b>11</b>	SISTEM INFORMASI SEKOLAH DALAM PENERAPAN SMART SCHOOL UNTUK MENINGKATKAN PELAYANAN SEKOLAH <b>Sofiansyah Fadli<sup>1</sup>, Ahmad Susan Pardiyansyah<sup>2</sup></b>	<b>95-108</b>

## ANALISIS SENTIMEN ULASAN EKSPEDISI J&T EXPRESS MENGGUNAKAN ALGORITMA NAÏVE BAYES

Mahardika Tania Nitami<sup>1</sup>, Herny Februariyanti<sup>2</sup>

<sup>1,2</sup>Program Studi Sistem Informasi Universitas Stikubank Semarang

Jln. Tri Lomba Juang No.1, Kec. Semarang Sel., Kota Semarang 50241

<sup>1</sup> [mahardikatan7@gmail.com](mailto:mahardikatan7@gmail.com), <sup>2</sup> [hernyfeb@edu.unisbank.ac.id](mailto:hernyfeb@edu.unisbank.ac.id)

---

### Abstract

*Sentiment analysis is a technique of identifying a person's opinion which is expressed in text form to analyze one's emotions related to a particular topic. J&T Express is a shipping service company with wide access in Indonesia and serves various shipments. In the era of globalization, expedition services are the services most widely used by the public. This is indicated by the increasing number of online purchases recommended by the government since the implementation of WFH. A large amount of use of the J&T Express service makes many users submit reviews regarding the performance of the delivery service. In this study, the sentiment analysis method is used to classify and process the text of J&T Express shipping service reviews. The data classification process in the sentiment analysis is divided into 2 classes, the negative sentiment and the positive sentiment. There are 500 data used and divided into 400 training data and 100 test data. Of the total 400 training data consisting of 237 negative classes and 163 positive classes. Based on the results of the research, it produces an accuracy of 87%, a positive class precision of 70%, a negative class precision of 95%, and an error rate of 13%.*

**Keywords :** *Naïve Bayes Classification, Sentiment Analysis, J&T Express, Text Mining*

### Abstrak

Analisis sentimen merupakan teknik mengidentifikasi pendapat seseorang yang diekspresikan dalam bentuk teks untuk menganalisa emosi seseorang terkait dengan topik tertentu. J&T Express menjadi sebuah perusahaan jasa pengiriman dengan akses luas di Indonesia serta melayani berbagai pengiriman. Pada era globalisasi, jasa ekspedisi menjadi jasa yang paling banyak digunakan oleh masyarakat. Hal ini ditandai dengan meningkatnya pembelian online yang direkomendasikan oleh pemerintah sejak diberlakukannya WFH. Besarnya penggunaan jasa J&T Express menjadikan banyak pengguna menyampaikan ulasan terkait kinerja jasa pengiriman tersebut. Didalam penelitian ini metode analisis sentimen digunakan untuk mengklasifikasi serta mengolah teks ulasan jasa pengiriman J&T Express. Proses klasifikasi data pada analisis sentimen dibagi 2 kelas, yaitu kelas sentimen negatif dan kelas sentimen positif. Terdapat 500 data yang digunakan dan terbagi menjadi 400 data latih dan 100 data uji. Dari total 400 data latih terdiri dari 237 kelas negatif dan 163 kelas positif. Berdasarkan hasil dari penelitian menghasilkan akurasi sebesar 87%, presisi kelas positif sebesar 70%, presisi kelas negatif sebesar 95%, dan error rate sebesar 13%.

**Kata kunci :** *Klasifikasi Naïve Bayes, Analisis Sentimen, J&T Express, Text Mining*

---

### 1. PENDAHULUAN

Perkembangan ilmu pengetahuan dan teknologi informasi pada masa ini sangat kencang dan pesat. Pertumbuhan kemajuan informasi membawa pengaruh baru yaitu membuka

peluang usaha untuk memasarkan produk secara global. Para pelaku bisnis memanfaatkan hal tersebut untuk mengembangkan bisnisnya melalui internet. Dan karenanya perdagangan elektronik menjadi suatu tempat atau zonayang



sangat menguntungkan bagi seseorang yang ingin ikut serta memulai masuk ke dalam dunia usaha perbisnisan melalui internet.[1]

Adanya pemasaran produk secara global tersebut menjadikan perkembangan ekspedisi barang juga mengalami kemajuan yang signifikan. Pada era globalisasi ini manusia dituntut harus mempunyai mobilitas tinggi. Selama adanya kegiatan pasar bebas di Indonesia, menjadikan tidak terbatasnya laju perdagangan barang serta jasa. Kebutuhan penggunaan jasa ekspedisi barang yang dipergunakan masyarakat untuk memenuhi berbagai kebutuhannya sangat meningkat pesat. Hadirnya berbagai jasa ekspedisi barang tidak hanya mempermudah masyarakat namun juga para pengusaha atau seller. Para pengusaha atau seller bisnis online ini menggunakan jasa ekspedisi untuk mengirimkan produk mereka kepada para pembeli yang tersebar di seluruh Indonesia hingga sampai ke luar negeri.[2]

Dari sekian banyak ekspedisi pengiriman di Indonesia salah satunya adalah J&T Express. J&T Express merupakan layanan jasa pengiriman yang dapat melayani berbagai pengiriman barang atau suatu produk keseluruh penjuru Indonesia, hal tersebut tentunya dapat mempermudah masyarakat dalam memenuhi kebutuhannya. Dengan banyaknya pengguna jasa pengiriman J&T Express menjadikan banyak pengguna yang menyampaikan opini-opini terkait jasa pengiriman tersebut dan dapat dimanfaatkan untuk menjadi sebuah informasi bagi yang membutuhkannya. Sejak masa pandemi kita diharuskan melakukan semua kegiatan di rumah untuk memutuskan rantai penularan virus corona. Mulai dari kegiatan pembelajaran dari rumah, bekerja dari rumah hingga berbelanja pun dianjurkan melakukan transaksi secara online. Dan karenanya jasa ekspedisi pengiriman tentunya sangat dibutuhkan untuk mengirim barang kepada pembelinya. Salah satu metode yang dapat digunakan untuk menganalisis opini-opini masyarakat tersebut yaitu analisa sentimen.

Analisa sentimen adalah proses penggunaan analisis teks untuk mendapatkan berbagai sumber data dari internet dan beragam platform media sosial. Analisis sentimen digunakan untuk menganalisa opini-opini masyarakat mengenai jasa pengiriman J&T Express. Pada penelitian ini penggunaan algoritma Naïve Bayes yaitu untuk mengetahui kecenderungan akan opini masyarakat yang berisi opini sentimen negatif dan sentimen positif berdasarkan dengan ulasan atau komentar yang disampaikan oleh para

pengguna jasa pengiriman J&T Express. Serta untuk melihat berapa besar nilai akurasi yang diperoleh oleh algoritma Naïve Bayes dalam mengklasifikasikan data ulasan pengguna ekspedisi J&T Express. [3]

## **2. TINJAUAN PUSTAKA DAN TEORI**

### **2.1. Tinjauan Pustaka**

Dalam penelitian yang dilaksanakan oleh Yonathan Sari Mahardhika (2018) yang menganalisis perihal bagaimana pemerintahan Joko Widodo dengan menggunakan algoritma Naïve Bayes. Penelitian ini bertujuan untuk mengetahui sentimen masyarakat akan pemerintahan yang dipimpin oleh Presiden Joko Widodo. Penelitian menggunakan bahasa pemrograman R, penelitian tersebut menghasilkan akurasi yang cukup tinggi sebesar 97% dari total 300 data latih dan 100 data uji yang dipakai.[4]

Pada penelitian selanjutnya Ghulam Asrofi Buntoro (2017) yang membandingkan 2 metode pengklasifikasian untuk mencari manakah akurasi tertinggi diantara ke dua algoritma yang digunakanyaitu Support Vector Machine (SVM) dan Naïve Bayes Classifier (NBC). Selain penggunaan kedua algoritma tersebut untuk menghitung nilai akurasi juga untuk mengetahui masalah opini masyarakat terhadap calon gubernur DKI Jakarta pada tahun 2017 yang memakai data tweet Twitter dengan 300 dataset. Hasil penelitian ini membuktikan bahwa metode klasifikasi Naïve Bayes mendapat nilai akurasi lebih tinggi dari metode Support Vector Machine. Yang memperoleh nilai akurasi sebesar 95%, presisi 95%, recall 95%.[5]

Di penelitian berikutnya Zia Ayu Nuansa Gumilang (2018) yang menerapkan metode Naïve Bayes Classifier dan Asosiasi untuk mengklasifikasikan sentimen negatif dan positif serta untuk mengetahui informasi yang banyak dibicarakan oleh pengguna pada aplikasi e-commerce Shopee. Hasil dari penelitian metode Naïve Bayes memperoleh akurasi sebesar 97,4%. Didapat asosiasi teks kelas positif yang paling banyak dibicarakan pengguna Shopee adalah aplikasi, ongkir, harga, dan puas. Asosiasi teks negatifnya adalah aplikasi ongkir, buruk, susah, gambar, pengiriman, dan penjual.[6]

Penelitian yang dilakukan oleh Riri Nada Devita, Heru Wahyu Herwanto, dan Aji Prasetya Wibawa (2018) mengenai perbandingan kinerja algoritma K-Nearest Neighbor dan Naïve Bayes untuk mengklasifikasi sekumpulan artikel dalam



Bahasa Indonesia. Kedua metode klasifikasi tersebut menunjukkan bahwa metode Naïve Bayes memiliki kinerja yang lebih unggul dibandingkan metode K-Nearest Neighbor. Nilai akurasi pada Naïve Bayes sebesar 70% dan K-Nearest Neighbor sebesar 40%.[7]

Pada penelitian berikutnya yang dilakukan oleh Eko Budi Santoso dan Aryo Nugroho (2019) perihal komentar publik di Facebook terhadap calon Presiden Indonesia 2019 didapati menggunakan metode Naïve Bayes Classifier dan teks asosisasi. Hasil penelitian yang menggunakan 5000 data tweet yang dipilih secara acak ini untuk calon presiden Joko Widodo yaitu mendapat 85% sentimen positif, 15% sentimen negatif sedangkan untuk Prabowo Subiyanto mendapat 76% sentimen positif, 24% sentimen negatif serta untuk nilai akurasi yang diperoleh sebanyak 86,4%.[8]

### 2.2. Text Mining

Text mining atau penambangan teks diartikan sebagai proses ekstraksi pola berupa informasi sebuah data besar yang tidak berstruktur dan bentuknya berupa teks. Penambangan teks terdapat dua tahapan proses yang bermuladari penerapan struktur yang berkenaan dengan sumber data teks, selanjutnya mengekstraksi informasi dan pengetahuan yang relevan dari data teks terstruktur dengan teknik dan alat yang serupa dengan penambangan data. Proses yang lazim digunakan oleh penambangan teks adalah perangkuman otomatis, kategorisasi dokumen, penggugusan teks, deteksi plagiarisme, dll. (Turban, et.al., 2011)[9]

### 2.3. Analisis Sentimen

Analisis sentimen adalah proses mengekstrak informasi yang didapat dari berbagai sumber data seperti internet dan beragam platform media sosial yang berbentuk pandangan seseorang terhadap suatu isu. Analisis sentimen juga dapat berupa opini kepuasan pelayanan, kebijakan. Dengan bantuan dari analisis sentimen, informasi yang pada awalnya tidak terstruktur diubah menjadi data yang lebih terstruktur.

### 2.4. Klasifikasi

Klasifikasi adalah sebuah proses untuk mengelompokkan dan memprediksi suatu data baru dari sebuah atribut atau variabel berdasarkan standar dari data yang sudah ada. Klasifikasi memiliki dua jenis data yaitu data uji

(testing data) dan data latih (training data). Data latih adalah data yang memang ada sebelumnya berdasarkan dari fakta yang telah terjadi contohnya data positif dan data negatif, juga berfungsi untuk membentuk tabel probabilitas. Sementara itu data uji adalah data yang telah berkelas atau berlabel dan berfungsi untuk menguji atau menghitung nilai akurasi dari data atau tabel probabilitas yang telah dibentuk.

### 2.5. Naïve Bayes Classifier

Naïve Bayes Classifier merupakan sebuah metode klasifikasi yang menggunakan metode probabilitas dan statistik. Algoritma naïve bayes digunakan untuk memprediksi peluang yang akan terjadi di masa depan berdasarkan dengan pengalaman dimasa sebelumnya. Naive bayes classifier memiliki ciri yaitu terdapat asumsi yang kuat (naif) terhadap independensi dari masing-masing kejadian. Algoritma ini memiliki 2 tahap dalam proses klasifikasi data, yaitu membuat data latih yang sudah berlabel positif dan negatif selanjutnya membuat data uji untuk mengetahui performa dari algoritma yang sudah dilatih sebelumnya. [10]

### 2.6. Confusion Matrix

Confusion matrix adalah metode yang digunakan untuk mengukur kinerja perhitungan akurasi pada sebuah metode klasifikasi data atau teks mining. Confusion matrix berisi informasi tentang klasifikasi aktual dan telah diprediksi oleh sistem klasifikasi (Kohavi & Provost, 1998).[11] Kinerja sistem yang dievaluasi biasanya menggunakan data dalam matriks. Tabel 2.1 dibawah ini menunjukkan confusion matrix dalam klasifikasi dua kelas.

Tabel 1 Confusion Matrix

Klasifikasi Asli	Klasifikasi	
	+	-
+	True Positive	False Negative
-	False Positive	True Negative

True positive adalah data positif yang diklasifikasi menjadi positif, false positive adalah data negatif yang diklasifikasi menjadi positif, false negative adalah data positif yang diklasifikasi menjadi negatif, true negative adalah data negatif yang diklasifikasi menjadi negatif. Setelah data uji dimasukkan dalam tabel confusion matrix, selanjutnya disusun dengan

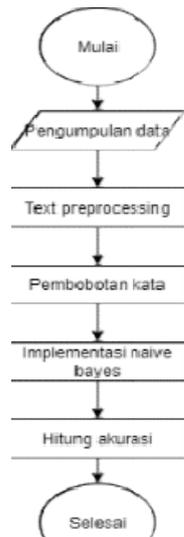
menghitung nilai-nilai data untuk dihitung jumlah sensitivity (recall), specificity, precision, dan accuracy.[12]

### 3. METODOLOGI PENELITIAN

#### 3.1. Skema Alur Penelitian

Analisis sentimen merupakan bidang studi yang menganalisis pendapat, sentimen, sikap, penilaian, evaluasi, serta emosi seseorang mengenai layanan, produk barang, topik, sesosok individu, dan kegiatan tertentu (Liu, 2012).[13] Analisis sentimen bertujuan untuk menentukan apakah opini suatu permasalahan dalam masyarakat cenderung kearah negatif atau positif serta analisis sentimen sering juga menjadi patokan untuk dapat meningkatkan kualitas sebuah barang atau meningkatkan kualitas suatu pelayanan jasa.[14]

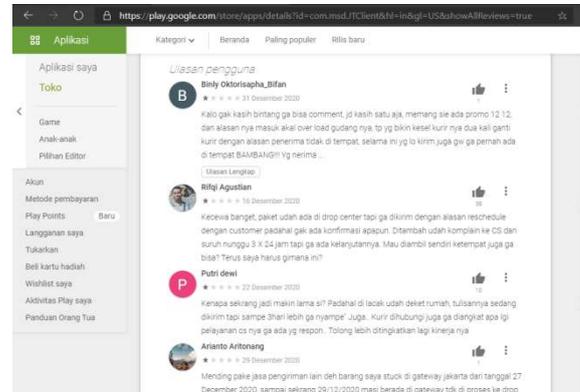
Pada gambar 1 menunjukkan skema alur penelitian yang dilakukan.



Gambar 1 Alur Penelitian

#### 3.2. Pengumpulan Data

Untuk proses pengumpulan data dalam penelitian yaitu data diambil secara manual dengan mengcopy-paste satu persatu data komentar para pengguna, pengumpulan data dapat juga dengan scraping data menggunakan aplikasi web scraping. Data yang terkumpul berjumlah 500 dataset dan data yang diambil adalah data komentar selama bulan Juli sampai dengan bulan September 2020. Gambar 2 menunjukkan proses dari pencarian data



Gambar 2 Data Ulasan

Kemudian 500 data ulasan tersebut disalin dan disimpan ke dalam microsoft excel dengan format xlsx, gambar 3 menunjukkan data ulasan yang sudah tersimpan.

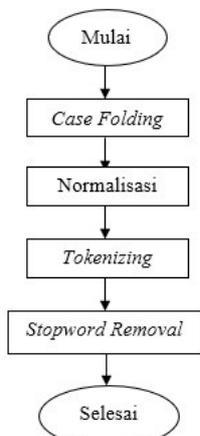
A	B	C	D
	class	text	tanggal
1	meg	Sumpah ini ekspedisi pengiriman yang paling telet, barang nyampe paling cepat 6 hari seringnya lambat	31/07/2020
2	meg	Pengiriman lambat sekali padahal dekat parah1	31/07/2020
3	meg	Ekspedisi gk punya integritas ekspedisi terburuk yg pernah tercipta jelek kali kek mana pun bagus jne	31/07/2020
4	pos	KAPOK PAKE JINT POKOKNYA KAPOK	31/07/2020
5	pos	Emang ekspres bgt....TOP	30/07/2020
6	pos	Makasih untuk aplikasi nya sangat bagus dan bermanfaat	30/07/2020
7	meg	Kurir j&t udah nelpon gw klo ada paket trus katanya ntar dia nelpon klo udah dekat rumah ehhh gw nunggu	30/07/2020
8	meg	seperti biasa paket tertunda entah apa alasannya ples deh jnt kalo memperkerkin kurir tuh yg bagus kimeri	19/09/2020
9	meg	Tidak ada fitur chat,pengiriman lamban meski barang sudah ada keterangan d antar kurtr... Layanan aneh beg	30/07/2020
10	meg	Tolong barangnya di kirim cepat,, masa ongkir uda mahal tapi lambat banget pengirimannya	30/07/2020
11	meg	Ahkir2 lambat	30/07/2020
12	pos	Gue sih heran bacain komentar nya.gue gopernah ada masalah sama jnt,kurir ramah, pengiriman cepat,ya	30/07/2020
13	pos	Pelayanan jnt benar benar buruk, saya pesan olshop pada tgl 23 yg satu memakai jne cashless yg satu jnt ex	30/07/2020
14	meg	Pelayanan Sagat lambat tolong diperbaiki...	30/07/2020
15	meg	Pengiriman lambat, respon lambat	30/07/2020
16	meg	tolong dengarkan keluh kesah para pelanggan anda, saya di sini juga termasuk salah satu pelanggan dan pe	30/07/2020
17	pos	Mantap	30/07/2020
18	meg	Jnt parah 6 paket gak jalan2 di gudang gak di kirim? gak amarah amat Pikir2 yang mau pake jasa pengirim	30/07/2020
19	pos	Pengiriman cepat..Abang kurir nya ramah,sudah langganan dengan j&t,	30/07/2020
20	nee	olexanan buruk	30/07/2020

Gambar 3 Data Ulasan Yang Tersimpan

#### 3.3. Text Preprocessing

Text processing merupakan tahap pertama dalam proses text mining dan tujuannya untuk membuat teks membentuk data untuk menjalani pengolahan ditahapan setelahnya.Pada tahap ini ada beberapa contoh kegiatan yang bisa dilakukan, dari yang bersifat sederhana misalnya parsing mengenai teks (pemecahan kalimat menjadi sekumpulan kata), dan dari yang bersifat kompleks misalnya partofspeech (pos) tagging, dan parse tree.[15]

Ketidakteraturan struktur data pada teks mining memerlukan proses awal yaitu mempersiapkan teks agar dapat diubah menjadi data yang terstruktur. Tahapan yang terdapat pada implementasi text mining yaitu tahap Text Preprocessing.Text preprocessing merupakan tahap dimana terjadinya proses seleksi data agar data yang didapat menjadi lebih terstruktur dalam setiap dokumen. Tahapan text preprocessing ditunjukkan pada gambar 4 dibawah ini.



Gambar 4 Alur Text Preprocessing

Pada tahapan text preprocessing terdapat 4 tahap proses yaitu:

1. Case Folding, yaitu proses untuk mengganti semua huruf di dalam data ke huruf kecil (lower case).
2. Normalisasi, yaitu proses untuk menghapus tanda baca, angka, dan juga karakter khusus pada teks atau dokumen.
3. Tokenizing, yaitu proses pemisahan sebuah kalimat menjadi satu kata dan lazimnya dipisahkan oleh karakter seperti tanda baca spasi, titik(.), dan koma(,).
4. Stopword Removal, yaitu tahap pemilahan kata yang bernilai dari hasil token dengan menyimpan kata yang bernilai dan membuang kata yang tidak bernilai.

### 3.4. Term Frequency

Term Frequency adalah frekuensi dari besarnya kemunculan kata pada sebuah dokumen. Semakin tinggi jumlah kemunculan suatu kata pada sebuah dokumen, maka semakin besar juga bobotnya dan hal tersebut menjadikan nilai kesesuaian akan menjadi semakin besar. Sebagai contoh jika muncul sebanyak empat (4) kali, maka kata atau term tersebut bernilai empat (4). Tabel 2 menunjukkan contoh hasil pembentukan term frequency dari salah satu data ulasan.

Tabel 2 Hasil Pembentukan Term Frequency

Term	Frequency
kualitas	1
pengiriman	1

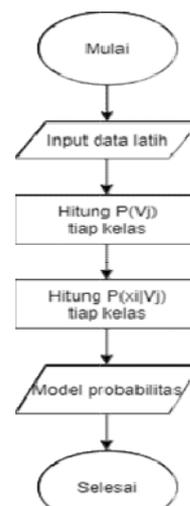
buruk	1
barang	1
rusak	1

### 3.5. Implementasi Algoritma Naïve Bayes Classifier

Metode pengklasifikasian dengan naïve bayes classifier menggunakan dua proses dalam pengklasifikasiannya yang dipergunakan untuk mendapatkan data latih dan data uji, diantaranya:

1. Proses Pelatihan Naïve Bayes Classifier

Didalam proses pelatihan naïve bayes memiliki beberapa tahapan, gambar 5 menunjukkan tahapan proses pelatihan naïve bayes.



Gambar 5 Proses Pelatihan Naïve Bayes

Naive bayes bisa dikatakan sebagai pembelajaran supervised, karenanya membutuhkan data awal pelatihan pada tahap pembelajarannya untuk dapat mengambil keputusan. Naive bayes menggambarkan sebuah perhitungan paling sederhana dari teorema bayes, karena sanggup menurunkan kompleksitas komputasi dan membentuk multiplikasi sederhana dari probabilitas. [16]

Sebelum memulai perhitungan naïve bayes terdapat beberapa hal yang dilakukan, yaitu menentukan nilai probabilitas dokumen terhadap kelas kata, dengan menggunakan persamaan dibawah ini.

$$P(V_j) = \frac{|docs\ j|}{|lat\ |} \tag{3.1}$$

Keterangan:

$P(V_j)$  = Probabilitas dari  $V_j$

$|doc\ j|$  = Total dari dokumen kategori  $j$

$|latih|$  = Total dari dokumen yang digunakan

untuk pelatihan

Yang dilakukan selanjutnya yaitu menentukan nilai probabilitas kosakata pada sebuah dokumen berdasarkan nilai probabilitas dokumen terhadap kelas dokumennya dengan menggunakan persamaan dibawah ini:

$$P(x_i|V_j) = \frac{nk+1}{n+|kosakata|} \quad (3.2)$$

Keterangan:

$P(x_i|V_j)$  = Probabilitas xi pada kategori  $V_j$

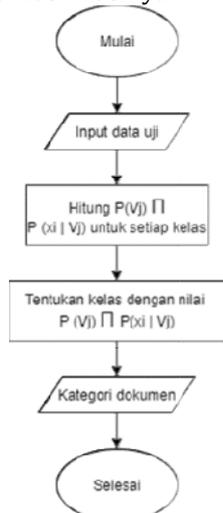
$nk$  = Total frekuensi kemunculan xi di kategori  $V_j$

$n$  = Total keseluruhan kata didokumen  $V_j$

$kosakata$  = Total kata didalam data latih

## 2. Proses Klasifikasi Naïve Bayes Classifier

Di dalam klasifikasi naïve bayes, data uji perlu melewati proses klasifikasi berdasarkan dengan data latih. Gambar 6 berikut ini adalah tahapan proses klasifikasinya:



Gambar 6 Proses Klasifikasi Naïve Bayes

Untuk melakukan perhitungan klasifikasi naïve bayes diambil dari data uji yang digunakan setelah dilakukannya tahapan text preprocessing. Selanjutnya, tahapan klasifikasi naïve bayes dimulai dengan menghitung nilai  $V_{map}$  pada tiap kelas dengan menggunakan persamaan dibawah ini

$$V_{map} = \underset{V_j \in V}{argmax} P(V_j) \prod P(x_i|V_j) \quad (3.3)$$

Keterangan:

$V_j$  = Kategori kelas kata

$P(V_j)$  = Probabilitas dari  $V_j$

$P(x_i|V_j)$  = Probabilitas xi di kategori  $V_j$

$V$  = Kategori kelas

$argmax$  = Titik maksimum probabilitas

## 3.6. Hitung Akurasi

Perhitungan akurasi hasil prediksi kelas sentimen pada klasifikasi naïve bayes, dihitung dengan menggunakan persamaan dibawah ini

$$Akurasi = \frac{TP+TN}{TP+FN+TN+FP} \quad (3.4)$$

Keterangan:

TP = True Positive

TN = True Negative

FN = False Negative

FP = False Positive

Data hasil klasifikasi dihitung berdasarkan banyaknya data benar dan banyaknya data yang salah. Dibawah ini diberikan contoh dari asumsi hasil klasifikasi dengan data uji sebanyak 50 data yang ditunjukkan oleh tabel confusion matrix.

$$Akurasi = \frac{TP+TN}{TP+FN+TN+FP} = \frac{24+16}{24+2+16+8} = \frac{40}{50} = 0,8$$

Tabel 3 dibawah merupakan hasil dari akurasi dan confusion matrix yang didapatkan dalam penelitian

Tabel 3 Hasil Confusion Matrix

		Actual	
		Positif	Negatif
Predict Class	Positif	24	2
	Negatif	8	16
Akurasi		80%	

## 4. HASIL DAN PEMBAHASAN

Didalam penelitian yang dilakukan menghasilkan klasifikasi data ulasan pengguna jasa ekspedisi J&T Express. Tahapan yang dilalui untuk melakukan proses klasifikasi yaitu data ulasan patut melewati beberapa tahapan pemrosesan. Untuk menjalankan pemrosesan klasifikasi pada penelitian ini menggunakan sebuah tools yaitu Rstudio untuk bahasa R.

### 4.1 Pembuatan Corpus

Corpus digunakan untuk mewakili setiap kata dalam dokumen sebagai token dan untuk setiap dokumen sebagai vektor fitur. Corpus tidak terfokus kepada urutan kata tetapi fokus pada jumlah kemunculan kata. Untuk membuat corpus menggunakan fungsi `Corpus(VectorSource())`. Hasil pembuatan corpus yang telah dibuat oleh RStudio dapat dilihat pada gambar 7 dibawah



Name	Type	Value
corpus	list [500] (S3: SimpleCorpus, Cor)	List of length 500
1	list [2] (S3: PlainTextDocument, T)	List of length 2
2	list [2] (S3: PlainTextDocument, T)	List of length 2
3	list [2] (S3: PlainTextDocument, T)	List of length 2
4	list [2] (S3: PlainTextDocument, T)	List of length 2
5	list [2] (S3: PlainTextDocument, T)	List of length 2
6	list [2] (S3: PlainTextDocument, T)	List of length 2
7	list [2] (S3: PlainTextDocument, T)	List of length 2
8	list [2] (S3: PlainTextDocument, T)	List of length 2
9	list [2] (S3: PlainTextDocument, T)	List of length 2
10	list [2] (S3: PlainTextDocument, T)	List of length 2
11	list [2] (S3: PlainTextDocument, T)	List of length 2
12	list [2] (S3: PlainTextDocument, T)	List of length 2

Gambar 7 Hasil Corpus

### 4.2 Text Preprocessing

Tahapan dalam text preprocessing meliputi case folding, normalisasi, tokenisasi, dan stopword removal. Dalam tahapan text preprocessing tersebut dapat dilakukan dengan memanggil library tm pada Rstudio. Sebelum teks diproses pada tahapan stopword removal, harus menginputkan data list stopwords. Stopword list yang digunakan adalah stopwords Tala Bahasa Indonesia dengan format csv. Data stopwords dapat dilihat pada gambar 8 dibawah

	A	B	C	D	E	F	G	H	I	J
1	ada									
2	adalah									
3	adanya									
4	adapun									
5	agak									
6	agaknya									
7	agar									
8	akan									
9	akankah									
10	akhir									
11	akhiri									
12	akhirnya									
13	aku									
14	akulah									
15	amat									
16	amatlah									
17	anda									
18	andalah									
19	antar									
20	antara									
21	antaranya									

Gambar 8 List Stopword Tala

Gambar 9 dan 10 merupakan hasil dari tahapan text preprocessing yang diantaranya

berupa case folding, normalisasi, tokenizing, dan stopword removal.

```
> inspect(corpus.clean)
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content: documents: 500

[1] alhamdulillah sesuai yg diharapkan semoga jaya trus mhn layanan kilat sehari

[2] sampah

[3] tolong perffisional gjl's bngt beda sma pelayanan jne

[4] jt buruk jelek kecewa

[5] cepat

[6] parah kurir jt kurir bernama gilang prihartono tasik parah paket ditungguin pagi sore ehh
anterin bikin status palsu kurir goblok

[7] salah sortir dll

[8] parah layanan pick transaksi bukalapaktokopedia respon
```

Gambar 9 Hasil Case Folding, Normalisasi, dan Stopword

```
dtm[["dimnames"]][["Terms"]]
[1] "alhamdulillah" "diharapkan"
[3] "jaya" "kilat"
[5] "layanan" "mhn"
[7] "sehari" "semoga"
[9] "sesuai" "trus"
[11] "sampah" "beda"
[13] "bngt" "gjl's"
[15] "jne" "pelayanan"
[17] "perffisional" "sma"
[19] "tolong" "buruk"
[21] "jelek" "kecewa"
[23] "cepat" "anterin"
[25] "bernama" "bikin"
[27] "ditungguin" "ehh"
[29] "gilang" "goblok"
[31] "kurir" "pagi"
[33] "paket" "palsu"
[35] "parah" "prihartono"
[37] "sore" "status"
[39] "tasik" "dll"
[41] "salah" "sortir"
[43] "bukalapakokopedia" "pick"
[45] "respon" "transaksi"
[47] "aduh" "judesnya"
[49] "nyesel" "pelayanannya"
[51] "hanus" "cepat"
```

Gambar 10 Hasil Tokenizing

### 4.3 Akurasi

Berdasarkan hasil prediksi kelas sentimen pada klasifikasi naïve bayes selanjutnya akan dihitung berapa hasil akurasi algoritma naïve bayes dalam mengklasifikasi sebuah data ulasan pengguna ekspedisi J&T Express. Gambar 11 merupakan hasil dari prediksi tabel pada 100 data uji yang diperoleh dari penelitian.

```
> summary(pred)
neg pos
73 27
```

Gambar 11 Prediksi Data Uji

Dapat dilihat pada gambar 11 diatas adalah prediksi dari 100 data uji yang digunakan dalam penelitian, dengan hasil prediksi yaitu 73 berlabel negatif dan 27 berlabel positif.

Hasil dari prediksi data uji algoritma naïve bayes menghasilkan tabel confusion matrix dengan pembagian data prediksi True Positif sebanyak 24 dokumen, True Negatif sebanyak 63 dokumen, False Positif sebanyak 10 dokumen, dan False Negatif sebanyak 3 dokumen. Dari data





	word	freq
barang	barang	133
pengiriman	pengiriman	122
paket	paket	119
jnt	jnt	119
kurir	kurir	116
kecewa	kecewa	71
cepat	cepat	63
bagus	bagus	57
pelayanan	pelayanan	56
buruk	buruk	44

Gambar 15 Data Frekuensi

## 5. Kesimpulan dan Saran

Berdasarkan hasil dari penelitian yang telah dilakukan dapat disimpulkan bahwa hasil dari pengujian klasifikasi naïve bayes pada data ulasan pengguna J&T Express dari bulan Juli sampai bulan September 2020 dapat dinyatakan bahwa ulasan cenderung termasuk kedalam kelas sentimen negatif. Algoritma Naïve Bayes Classifier teruji efektif dan dalam mengklasifikasikan data ulasannya cukup menjamin karena menghasilkan nilai akurasi yang cukup tinggi yaitu 87%.

Untuk penelitian selanjutnya, dapat dilakukan dengan menggunakan algoritma selain Naïve Bayes Classifier seperti Support Vector Machine (SVM) atau algoritma K-Nearest Neighbor (KNN), untuk memahami manakah algoritma yang menghasilkan hasil terbaik dalam mengklasifikasikan sebuah data ulasan. Serta memperbanyak data yang digunakan untuk mengklasifikasi agar didapatkan hasil yang akurat dan lebih baik.

## 6. UCAPAN TERIMA KASIH

Saya kirimkan ucapan terima kasih kepada Universitas Stikubank Semarang, Program Studi Sistem Informasi, dan Ibu Herny Februariyanti, ST, M.Cs yang membantu memberikan masukan, motivasi, serta dukungan hingga penyusunan tulisan artikel ini dapat teratasi sampai akhir.

## Daftar Pustaka:

- [1] H. Nurul, "Analisis Kepuasan Konsumen Jasa Pengiriman Barang J&T Ekspres ditinjau dari Kualitas Pelayanan, Fasilitas, dan Harga," *Article*, p. 13, 2018.
- [2] A. F. Kamal, "Text Mining untuk Analisa Sentiment Ekspedisi Jasa Pengiriman Barang Menggunakan Metode Naive Bayes pada Aplikasi J&T Express," 2017.
- [3] N. I. P. Kalingara, O. N. Pratiwi, and H. D. Anggana, "Analisis Sentimen Review Customer Terhadap Layanan Ekspedisi Jne Dan J & T Express Menggunakan Metode Naïve Bayes," vol. 8, no. 5, pp. 9035-9048, 2021.
- [4] Y. S. Mahardhika and E. Zuliarso, "Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes," *Pros. SINTAK 2018*, no. 2015, pp. 409-413, 2018.
- [5] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di Twitter," *Integer J. Maret*, vol. 1, no. 1, pp. 32-41, 2017, [Online]. Available: [https://www.researchgate.net/profile/G\\_hulam\\_Buntoro/publication/316617194\\_Analisis\\_Sentimen\\_Calon\\_Gubernur\\_DKI\\_Jakarta\\_2017\\_Di\\_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf](https://www.researchgate.net/profile/G_hulam_Buntoro/publication/316617194_Analisis_Sentimen_Calon_Gubernur_DKI_Jakarta_2017_Di_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf).
- [6] Z. A. N. Gumilang, "Implementasi Naive Bayes Classifier dan Asosiasi untuk Analisis Sentimen Data Ulasan Aplikasi E-Commerce Shopee pada Situs Google Play," Islamic University of Indonesia, 2018.
- [7] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, "Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, p. 427, 2018, doi: 10.25126/jtiik.201854773.
- [8] E. B. Santoso and A. Nugroho, "Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik di Facebook," *Eksplora Inform.*, vol. 9, no. 1, pp. 60-69, 2019, doi: 10.30864/eksplora.v9i1.254.
- [9] E. . et. al. Turban, *Decision Support and Business Intelligence Systems. New Jersey: Pearson Education, Inc.* 2011.
- [10] O. Gunawan, "Analisis Sentimen Masyarakat Terhadap Layanan Pengaduan di Kota Semarang 'Lapor Hendi' Berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes," 2020.
- [11] R. Kohavi and F. Provost, *Glossary of terms. Machine Learning—Special Issue on Applications of Machine Learning and the Knowledge Discovery Process.*, vol. 30.



- 1998.
- [12] Saikin and Kusriani, "Model Data Mining Untuk Karakteristik Data Traveller Pada Perusahaan Tour and Travel," *J. Manaj. Inform. dan Sist. Inf.*, vol. 2, no. 2, p. 61, 2019, doi: 10.36595/misi.v2i2.105.
- [13] Liu B, "Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers," 2012.
- [14] W. E. Nurjanah, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter Menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 12, pp. 1750–1757, 2017, doi: 10.1074/jbc.M209498200.
- [15] I. W. D. Pancane and I. W. Suriana, "Klasifikasi Komentar Publik Terhadap Kebijakan Pemerintah Pada Facebook Frontpage Kompas Menggunakan Clustering K-Means, Furthest First," *J. Manaj. Inform. dan Sist. Inf.*, vol. 3, no. 2, 2020.
- [16] D. Sartika and D. Indra, "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian," *J. Tek. Inform. Dan Sist. Inf.*, vol. 1, no. 2, pp. 151–161, 2017.
- [17] Imtihan, K., Bagye, W., Asri, Z. M. T., Fadli, S., & Ashari, M. (2021, February). Image capture device based on Internet of Thing (IoT) technology. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1088, No. 1, p. 012065). IOP Publishing
- [18] Bagye, W., Imtihan, K., Ashari, M., Fadli, S., Zaen, M. T., Zulkarnaen, M. F., ... & Tantoni, A. (2021, February). The potential of hand tractor controller to reduce the risk of Hand-Arm Vibration Syndrome (HAVS). In *IOP Conference Series: Materials Science and Engineering* (Vol. 1088, No. 1, p. 012077). IOP Publishing.