

Implementasi Algoritma *K-Nearest Neighbor* dan *Probabilistic Neural Network* untuk Analisis Opini Masyarakat Terhadap Toko *Online* di Indonesia

Assad Hidayat¹, Mustakim⁴, Muhammad Zakiy Fauzi², Imaduddin Syukra³

^{1,2,3,4}Puzzle Research Data Technology (Predatech), Fakultas Sains dan Teknologi

Universitas Islam Negeri Sultan Syarif Kasim Riau

^{1,2,3,4}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi

Universitas Islam Negeri Sultan Syarif Kasim Riau

Jl. HR. Soebrantas No. 155 Simpang Baru, Tampan. Pekanbaru, Riau – Indonesia 28293

e-mail: hidayatassad@gmail.com, mustakim@uin-suska.ac.id,

zakiyfauzi@gmail.com, imaduddinsyukra@gmail.com

Abstrak

Twitter dipakai oleh semua orang untuk melakukan penilaian dan mengeluarkan pendapat mengenai segala sesuatu dan mengunggah postingan dengan opini yang berbeda-beda melalui tweet. Tweet yang digunakan berasal dari akun resmi toko online pada Twitter yaitu Lazada dan Blibli. Tweet tersebut dianalisis menggunakan sebuah teknik penggalian teks atau disebut juga text mining. Pada penelitian ini menggunakan dua algoritma klasifikasi yang terdapat pada text mining yaitu *K-Nearest Neighbor* dan *Probabilistic Neural Network*, dimana kelas data dibagi menjadi tiga yaitu netral, negatif dan positif yang ditentukan dengan cara menggunakan kamus sentimen. Data dibagi menggunakan *K-Fold Cross Validation* yang kemudian dihitung akurasi. Pada percobaan menggunakan metode pembagian data dengan 10 K pada *K-Fold Cross Validation* didapat hasil perbandingan akurasi antara KNN dan PNN terhadap data Lazada dan data Blibli, dimana pada data Lazada akurasi KNN lebih tinggi dibandingkan PNN dengan akurasi KNN sebesar 71.57% sedangkan PNN sebesar 66.71%. Untuk data Blibli akurasi KNN juga lebih tinggi dibandingkan PNN dengan akurasi KNN sebesar 68.29% sedangkan PNN sebesar 65.29%. Jadi hasil akurasi pada data Lazada dan Blibli menggunakan algoritma KNN memiliki performa lebih baik dibanding PNN.

Kata kunci: Akurasi, *K-Fold Cross Validation*, Klasifikasi, *k-Nearest Neighbor*, *Probabilistic Neural Network*, Text Mining, Twitter.

Abstract

Twitter is used by everyone to make judgments and express opinions about everything and upload posts with different opinions through tweets. Tweets used come from the official online store accounts on Twitter, Lazada, and Blibli. Tweets are analyzed using a technique of extracting text or also called text mining. In this research using two classification algorithms found in text mining, namely *K-Nearest Neighbor* and *Probabilistic Neural Network*, where the data class is divided into three, namely neutral, negative and positive, which is determined by using a sentiment dictionary. Data is processed using *K-Fold Cross Validation* which is then calculated for accuracy. In the experiment using 10 K data sharing method on *K-Fold Cross Validation* and obtained the results of the comparison of accuracy between KNN and PNN against Lazada data and Blibli data, Lazada data has higher KNN accuracy when compared to PNN, with an KNN accuracy of 71.57%, whereas PNN has an accuracy of 66.71%. For accurate Blibli data, KNN is also higher than PNN, with an accurate KNN of 68.29%, while PNN is 65.29%. So, the accuracy of Lazada and Blibli data using KNN algorithm has better performance than PNN.

Keywords: Accuracy, *K-Fold Cross Validation*, Classification, *k-Nearest Neighbor*, *Probabilistic Neural Network*, Text Mining, Twitter

1. Pendahuluan

Toko *online* dijadikan sebagai media dalam proses berbelanja secara langsung yang memanfaatkan fasilitas yang terdapat pada sebuah *website* yang menyediakan layanan pesanan dan pengiriman [1]. Berdasarkan hasil survei *iPrice E-commerce Merchants Award* (iEMA) 2018, toko *online* yang populer di Twitter adalah Blibli sebanyak 474.700 pengikut dan Lazada sebanyak 363.600 pengikut [2]. Akun resmi toko *online* tersebut terdapat ulasan-ulasan yang disampaikan melalui komentar yang berisi opini-opini yang diunggah oleh masyarakat terhadap isu-isu yang sedang terjadi. Seperti permasalahan di toko *online* yaitu pada standar logistik, hal ini menjadi masalah yang paling sering dikeluhkan oleh masyarakat sehingga memunculkan perilaku complain dan kurangnya kepercayaan masyarakat [3]. Suatu kepercayaan di toko *online* dalam bertransaksi sangatlah penting bagi masyarakat, hal ini menjadi salah satu faktor utama dalam kepuasan dan

niat masyarakat untuk berbelanja suatu produk di toko *online* [4]. Masalah-masalah tersebut dapat dianalisis menggunakan teknik *text mining* untuk mengetahui opini masyarakat terhadap toko *online*.

Text mining diartikan sebagai suatu cara pengambilan informasi dimana terjadinya suatu interaksi antara pengguna dengan kumpulan dokumen menggunakan alat analisis dari waktu ke waktu [5]. Penelitian *text mining* biasanya mengarah pada bidang *data mining*, karena *text mining* memiliki tingkat arsitektur yang sama dengan *data mining* [6]. Penelitian tentang *text mining* telah banyak diterapkan, diantaranya penelitian oleh Aulianita (2016), mengenai perbandingan algoritma *k-Nearest Neighbor* (KNN) dan *Support Vector Machine* (SVM) pada analisis sentimen review kamera, dengan hasil, algoritma KNN memiliki akurasi terbaik dibandingkan dengan algoritma SVM yaitu hasil KNN 79.0% sedangkan SVM 72.0% [7]. Selanjutnya, penelitian yang dilakukan oleh Alam dan Yao (2018) yang menggunakan *Probabilistic Neural Network* (PNN) dan pemodelan kata (*Word Embedding*) untuk analisis sentimen. Penelitian tersebut membandingkan hasil akurasi antara algoritma PNN dengan SVM, NB dan MaxE, dimana PNN memiliki akurasi tertinggi yaitu 98.0% [8]. Pada penelitian sebelumnya, *text mining* dapat diterapkan menggunakan teknik yang ada pada *data mining*, salah satunya teknik klasifikasi.

Teknik klasifikasi adalah suatu cara untuk mencari gambaran label kelas yang sedemikian rupa, sehingga dapat digunakan untuk memprediksi label kelas yang tidak diketahui [9]. Dengan demikian, klasifikasi biasanya digunakan untuk memprediksi label kelas yang belum diketahui. Proses klasifikasi melibatkan dua tahap, langkah awal ialah tahap pembelajaran yang melibatkan pembangunan model sedangkan langkah kedua melibatkan penggunaan model untuk memprediksi label kelas [10].

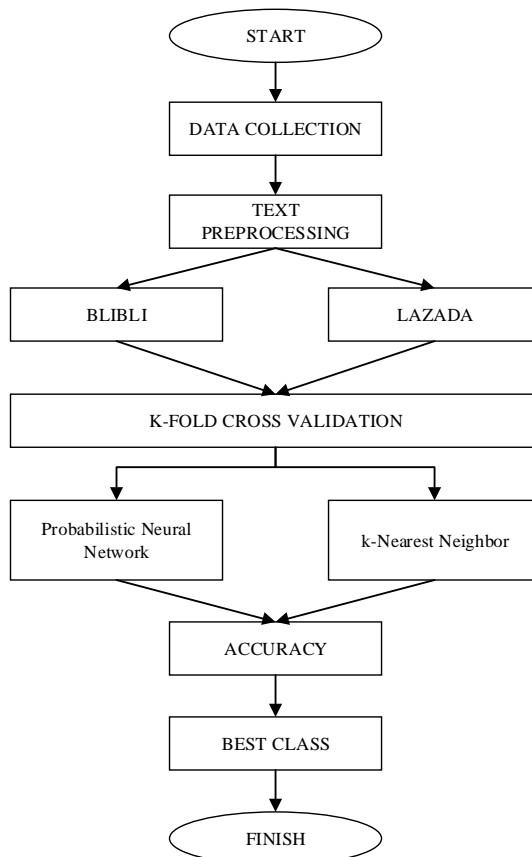
Penelitian ini menggunakan dua algoritma yang terdapat pada teknik klasifikasi yakni *k-Nearest Neighbor* (KNN) dan *Probabilistic Neural Network* (PNN). KNN adalah algoritma yang sederhana, dimana suatu objek diklasifikasikan berdasarkan pada banyak tetangganya [11]. Sedangkan algoritma PNN adalah suatu proses pada jaringan syaraf tiruan probabilistik yang termasuk dalam pembelajaran yang terawasi, karena sebelumnya keluaran yang dihasilkan telah diketahui [12]. Kelebihan yang ada pada algoritma PNN salah satunya adalah modifikasi jaringan yang dilakukan adanya kemudahan dalam penambahan atau pengurangan terhadap data latih yang digunakan [9]. Pada penelitian ini akan membandingkan antara algoritma KNN dan PNN untuk memperoleh hasil akurasi tertinggi dari kedua algoritma tersebut dan hasil klasifikasi diharapkan memberikan kemudahan dalam pembacaan data tentang opini masyarakat terkait toko *online* di Twitter.

2. Metode Penelitian

Adapun langkah-langkah yang ditempuh dalam penelitian ini terdapat pada Gambar 1.

2.1. *Text Mining* dan *Text Preprocessing*

Text mining biasanya disebut dengan *text data mining* yaitu mencari informasi berbasis data tekstual atau proses semi otomatis untuk melakukan ekstraksi pola dari sebuah data [14]. Menurut Feldman dan Sanger (2007) *text mining* ialah proses menemukan informasi dimana pengguna berinteraksi dengan sekumpulan dokumen dari waktu ke waktu dengan menggunakan seperangkat alat analisis. *Text mining* merupakan suatu metode yang dipakai untuk menyelesaikan masalah *information retrieval*, *information extraction*, *clustering* dan klasifikasi [15]. Pada *text mining* perlu adanya proses dalam mengolah data teks agar menjadi terstruktur. Tahapan *text pre-processing* adalah salah satu proses dalam *text mining* [16]. Tahap *text pre-processing* yang digunakan dalam penelitian ini adalah *cleaning*, *filtering*, *tokenizing* dan *stemming*.



Gambar 1. Metodologi Penelitian

3.1. Algoritma Nazief dan Adriani

Bobby Nazief dan Mirna Adriani mengembangkan sebuah algoritma yang dinamakan algoritma Nazief dan Adriani. Algoritma ini berlandaskan aturan morfologi bahasa Indonesia yang cakupannya luas, lalu dikumpulkan menjadi sebuah satu kesatuan dan di-enkapsulasi pada imbuhan kata yang dibolehkan dan yang tidak diperbolehkan. Algoritma ini menggunakan kata dasar yang terdapat pada kamus yang mendukung proses *recoding*, yaitu penyusunan kembali kata-kata yang mengalami proses *stemming* berlebih [17].

4.1. K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) merupakan algoritma yang biasanya digunakan untuk proses klasifikasi data. KNN merupakan suatu tahapan yang menggunakan algoritma terawasi, dimana hasil dari *query instance* yang baru nantinya diklasifikasikan berdasarkan mayoritas dari kategori yang ada [20]. Algoritma ini digunakan untuk mengklasifikasikan objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Dekat atau jauhnya jarak biasanya dihitung berdasarkan *Euclidean Distance* yang direpresentasikan pada persamaan berikut:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \quad (1.1)$$

Dimana $d(x_i, x_j)$ didapat dari pengurangan *record* (x_i) dari setiap atribut lalu di kuadratkan, kemudian hasil kuadrat dijumlahkan dan diurutkan berdasarkan nilai terkecil untuk melihat jarak terdekat dengan data uji [27].

5.1. Probabilistic Neural Network (PNN)

Jaringan Syaraf Tiruan (JST) merupakan proses *data modeling* yang mampu mewakili keterkaitan antara keluaran dan masukan yang rumit, karena memiliki kelebihan dalam menyelesaikan beberapa masalah yang relatif mudah untuk digunakan dan ketahanan dalam menginput data dengan cepat saat dieksekusi, dan dapat menginisialisasikan sistem yang kompleks

[21]. *Probabilistic Neural Network* (PNN) adalah salah satu bidang pada JST yang banyak dikembangkan dan dipakai orang-orang untuk menyelesaikan masalahnya. PNN memiliki waktu pelatihan yang relatif singkat dan memiliki tingkat akurasi yang lumayan tinggi. Salah satu faktor yang menyebabkan akurasinya menjadi cukup tinggi karena menggunakan *spread* [22].

$$p(x|C_k) = \frac{1}{(2\pi)^{\frac{m}{2}} \sigma_k^m} \sum_{p_i \in C_k} \exp[-||x - w_i||^2 / (2\sigma_k^2)] \quad (1.2)$$

Dimana C_k merupakan total pola latih pada kelas, m merupakan dimensi vektor pola masukan, w_i adalah vektor bobot pada pola latih ke- i . Dan σ_k merupakan parameter penghalus (*spread*).

6.1. Probabilistic Neural Network (PNN)

Toko *online* (*E-Commerce*) merupakan sebuah prosedur dimana penjual dan pembeli dipertemukan di dunia maya untuk melakukan perdagangan atau mekanisme jual-beli melalui internet. Toko *online* dapat juga dijadikan sebagai media dalam proses berbelanja secara langsung yang memanfaatkan fasilitas yang terdapat pada sebuah *website* yang menyediakan layanan pesanan dan pengiriman [1].

3. Hasil dan Pembahasan

3.1. Analisis Data

Data yang digunakan pada penelitian ini yaitu data yang berasal dari twitter, dengan kata kunci yang berhubungan dengan akun resmi toko *online* dengan rentang waktu januari 2018 sampai dengan september 2018. Total data keseluruhan sebanyak 19.990 data awal. Pengumpulan data merupakan tahapan yang dilakukan untuk mendapatkan data yang akan digunakan dalam penelitian. Pengumpulan data pada penelitian ini dilakukan dengan cara *crawling* pada Twitter menggunakan bahasa pemrograman Python. Pengumpulan data dilakukan dengan menggunakan kata kunci pencarian yang berhubungan dengan akun resmi toko online yang digunakan untuk melayani konsumen atau tanggapan pada Twitter.

Data awal berasal dari hasil *crawling* dengan kata kunci “@LazadaIDCare” dan “@BibliCare” yang langsung diambil dari Twitter dengan menggunakan bahasa pemrogram python. Penentuan atribut adalah tahapan menentukan atribut data yang akan digunakan sesuai dengan kebutuhan dan yang berhubungan dengan penelitian. Adapun atribut atau kriteria yang digunakan pada penelitian ini yaitu teks atau *text* dari data Twitter. Data teks yang tersedia pada Twitter adalah data yang tidak berlabel atau *unsupervised data*. Sehingga untuk diproses menggunakan *supervised data* diperlukan sebuah metode untuk melakukan pelabelan.. Kamus sentimen positif dan negatif yang digunakan berasal dari Wahid dan Azhari (2016) yang sudah dimodifikasi dan sudah diartikan ke bahasa indonesia dari versi aslinya yaitu Liu dkk.

3.2. K-Nearest Neighbor (KNN)

Penerapan algoritma KNN dalam penelitian ini digunakan untuk mendapatkan hasil akurasi dari klasifikasi *text mining* menggunakan data *tweet* yang berasal dari Twitter. KNN pada penelitian ini menggunakan rumus jarak *Euclidean Distance* yang terdapat pada persamaan 2.2, dan menggunakan enam parameter K yaitu $k = 5$, $k = 10$, $k = 25$, $k = 50$, $k = 100$ dan $k = 200$.

3.2.1. K-Fold Cross Validation KNN pada Data Lazada

Hasil percobaan pembagian data latih dan data uji menggunakan *K-Fold Cross Validation* pada KNN untuk data Lazada dapat dilihat pada Tabel 1.

Tabel 1. *K-Fold Cross Validation KNN* pada Data Lazada

K-Fold	Parameter K					
	K = 5	K = 10	K = 25	K = 50	K = 100	K = 200
1	63.71%	67.57 %	66.57 %	66.14 %	64.43 %	60.14 %
2	62.43 %	64.29 %	66.86 %	66.57 %	64.57 %	60.43 %
3	69.29 %	70.86 %	71.57 %	71.43 %	68.29 %	63.43 %
4	68.71 %	69.29 %	69.14 %	69.14 %	69.43 %	66.00 %
5	70.43 %	68.43 %	68.29 %	67.14 %	67.29 %	60.71 %

K-Fold	Parameter K					
	K = 5	K = 10	K = 25	K = 50	K = 100	K = 200
6	63.81 %	62.66 %	63.66 %	63.23 %	62.95 %	60.66 %
7	60.80 %	61.95 %	62.52 %	60.80 %	60.23 %	56.65 %
8	64.23 %	64.52 %	66.38 %	65.67 %	64.95 %	61.95 %
9	65.81 %	64.66 %	64.95 %	67.24 %	64.81 %	63.38 %
10	68.53 %	68.24 %	69.53 %	69.67 %	69.24 %	66.38 %
Nilai Max	70.43 %	70.86 %	71.57%	71.43 %	69.43 %	66.38 %

3.2.2. K-Fold Cross Validation KNN pada Data BliBli

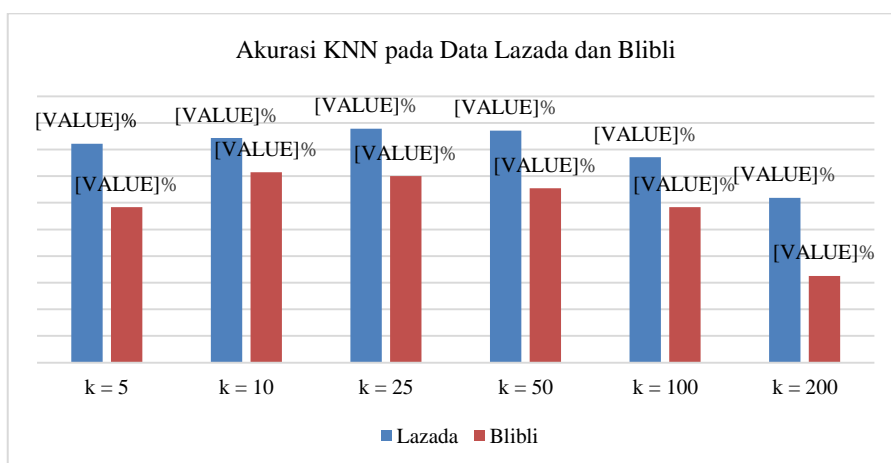
Hasil percobaan pembagian data latih dan data uji menggunakan *K-Fold Cross Validation* pada KNN untuk data BliBli dapat dilihat pada Tabel 2.

Tabel 2. *K-Fold Cross Validation* KNN pada Data BliBli

K-Fold	Parameter k					
	K = 5	K = 10	K = 25	K = 50	K = 100	K = 200
1	58.57 %	57.57 %	59.71 %	56.86 %	54.71 %	51.71 %
2	58.14 %	61.57 %	60.57 %	58.14 %	56.14 %	53.57 %
3	60.71 %	61.86 %	61.43 %	60.86 %	59.14 %	54.71 %
4	61.29 %	63.00 %	61.43 %	57.43 %	54.00 %	50.14 %
5	62.86 %	66.00 %	66.00 %	63.29 %	59.14 %	54.86 %
6	65.14 %	68.29 %	68.00 %	64.43 %	56.14 %	49.57 %
7	63.57 %	65.57 %	61.86 %	61.57 %	58.71 %	57.86 %
8	63.09 %	65.38 %	64.09 %	56.22 %	53.51 %	51.50 %
9	62.80 %	63.95 %	63.81 %	60.23 %	58.23 %	56.08 %
10	65.67 %	66.95 %	67.53 %	67.10 %	65.67 %	60.52 %
Nilai Max	65.67%	68.29%	68.00%	67.10%	65.67%	60.52%

3.2.3. Perbandingan Akurasi KNN pada Data Lazada dan BliBli

Perbandingan akurasi KNN antara Data Lazada dan data BliBli yang tertinggi adalah data Lazada dengan akurasi sebesar 71.57% pada parameter k = 25, sedangkan data BliBli sebesar 68.29% pada parameter k = 10. Hasil perbandingan akurasi KNN dengan menggunakan pembagian data *K-Fold Cross Validation* pada data Lazada dan data BliBli dapat dilihat pada Gambar 2.



Gambar 2. Perbandingan Akurasi KNN pada Data Lazada dan BliBli

3.3. Probabilistic Neural Network (PNN)

Penerapan Algoritma PNN dalam penelitian ini digunakan untuk mendapatkan hasil akurasi dari pengolahan data opini masyarakat tentang toko *online* pada Twitter dengan konsep klasifikasi. PNN memiliki arsitektur yang terdiri dari empat *layer*, yaitu *Input Layer*, *Pattern Layer*, *Summation Layer* dan *Decision Layer* atau *Output Layer*.

3.3.1 Probabilistic Neural Network (PNN)

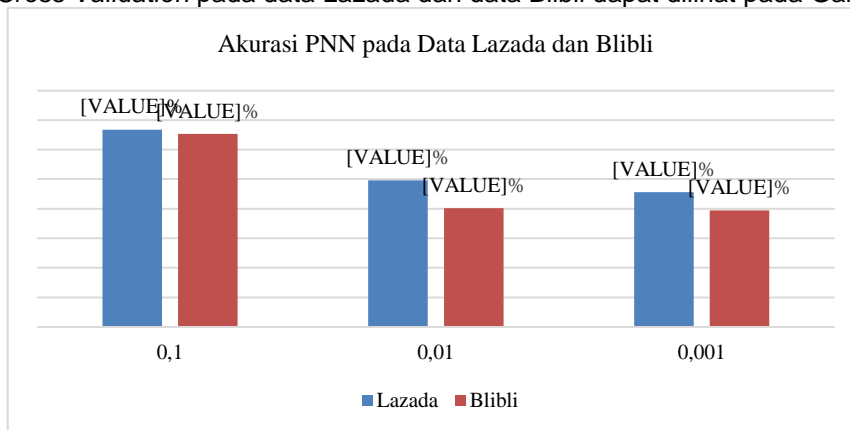
Hasil percobaan pembagian data latih dan data uji menggunakan *K-Fold Cross Validation* pada PNN untuk data Blibli dapat dilihat pada Tabel 4.

Tabel 4. *K-Fold Cross Validation* PNN pada Data Lazada

K-Fold	Nilai Spread		
	0.1	0.01	0.001
1	56.57%	34.43%	33.43%
2	55.43%	34.71%	34.14%
3	57.71%	36.57%	35.86%
4	57.14%	34.00%	33.57%
5	60.57%	31.86%	31.00%
6	63.14%	29.86%	28.43%
7	65.29%	40.14%	39.43%
8	59.23%	30.62%	28.04%
9	59.80%	36.05%	34.48%
10	60.66%	35.48%	34.62%
Nilai Max	65.29%	40.14%	39.43%

3.3.2 Perbandingan Akurasi KNN pada Data Lazada dan Blibli

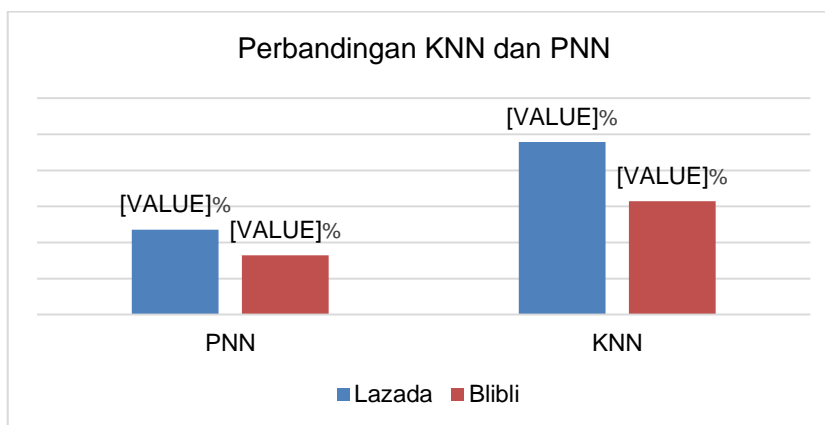
Perbandingan akurasi PNN antara Data Lazada dan data Blibli yang tertinggi adalah data Lazada dengan akurasi sebesar 66.71% pada nilai *spread* = 0.1, sedangkan data Blibli sebesar 65.29% pada *spread* = 0.1. Hasil perbandingan akurasi PNN dengan menggunakan pembagian data *K-Fold Cross Validation* pada data Lazada dan data Blibli dapat dilihat pada Gambar 3.



Gambar 3. Perbandingan Akurasi PNN pada Data Lazada dan Blibli

3.3.2 Analisis Klasifikasi

Pada percobaan menggunakan metode pembagian data dengan *K-Fold Cross Validation* didapat perbandingan akurasi antara KNN dan PNN terhadap data Lazada dan data Blibli, Dimana pada data Lazada akurasi KNN lebih tinggi dibandingkan PNN dengan akurasi KNN sebesar 71.57% sedangkan PNN sebesar 66.71%. Untuk data Blibli akurasi KNN juga lebih tinggi dibandingkan PNN dengan akurasi KNN sebesar 68.29% sedangkan PNN sebesar 65.29%. Jadi hasil akurasi pada data Lazada dan Blibli menggunakan algoritma KNN lebih unggul dibandingkan PNN. Hasil perbandingan akurasi KNN dan PNN pada data Lazada dan Blibli dapat dilihat pada Gambar 4.



Gambar 4. Perbandingan Akurasi PNN dan KNN

4. Kesimpulan

Pada percobaan menggunakan metode pembagian data dengan 10 K pada *K-Fold Cross Validation* didapat hasil perbandingan akurasi antara KNN dan PNN terhadap data Lazada dan data Blibli, Dimana pada data Lazada akurasi KNN lebih tinggi dibandingkan PNN dengan akurasi KNN sebesar 71.57% sedangkan PNN sebesar 66.71%. Untuk data Blibli akurasi KNN juga lebih tinggi dibandingkan PNN dengan akurasi KNN sebesar 68.29% sedangkan PNN sebesar 65.29%. Jadi hasil akurasi pada data Lazada dan Blibli menggunakan algoritma KNN memiliki performa lebih baik dibanding PNN.

Daftar Pustaka

- [1] Aryato, D., Tarigan, F.A., "Aplikasi Penjualan Pakaian Secara Online". Jurnal TIMES, Vol. 4, No. 1, 2015.
- [2] iPrice Indonesia. *Peta E-commerce Indonesia*. Juli 2018.
<https://iprice.co.id/insights/mapofecommerce/>, diakses 8 September, 2018
- [3] Putri, R.O., Wibawa, B.M., Laksamana, T. "Identifikasi Permasalahan Komplain pada E-Commerce Menggunakan Metode Fishbone" Jurnal Sains dan Seni ITS, Vol. 6 No. 1 Hal. 37-41, 2017
- [4] Andhini, A., "Pengaruh Transaksi Online Shopping, Dan Kepercayaan Konsumen Terhadap Kepuasan Konsumen Pada E-Commerce" Jurnal Ilmu dan Riset Manajemen, Vol. 6, No. 7, 2017.
- [5] Falahah., Nur, D.D.A., "Pengembangan Aplikasi Sentiment Analysis Menggunakan Metode Naïve Bayes" Seminar Nasional Sistem Informasi Indonesia, November 2015.
- [6] Feldman, R., dan J. Sanger. (2007). "The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data". Cambridge University Press. New York.
- [7] Aulianita, R., "Komparasi Metode k-Nearest Neighbors dan Support Vector Machine Pada Sentiment Analysis Review Kamera" Jurnal Speed, Vol. 8, No. 3, 2016
- [8] Alam, S., Yao, N., "Probabilistic Neural Network and Word Embedding for Sentiment Analysis" International Journal of Advanced Computer Science and Applications, Vol. 9, No. 7, 2018.
- [9] Hidayat, A., Effendi, Z., Aszani., Novita, R., Lestari, T.L. "Algorithm Comparison Of Naive Bayes Classifier And Probabilistic Neural Network For Water Area Classification Of Fishing Vessel In Indonesia". Journal of Theoretical and Applied Information Technology, Vol. 96 No. 13 Hal 4114-4125, Juli 2018.
- [10] Singh, A., Sathyaraj, R., "A Comparison Between Classification Algorithms on Different Datasets Methodologies using Rapidminer", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Issue 5, May 2016
- [11] Padmaja, S., & Fatima, S.S. "Opinion Mining and Sentiment Analysis –An Assessment of Peoples' Belief: A Survey" International Journal of Ad hoc, Sensor & Ubiquitous Computing, Vol. 4, No. 1, February 2013.

- [12] Lotfi, A., & Benyettou, A. "A Reduced Probabilistic Neural Network For The Classification Of Large Databases", Turkish Journal Of Electrical Engineering & Computer Sciences. Doi:10.3906/Elk-1209-35 Algeria. 2014
- [13] Gorunescu, F. (2011). *Data Mining : Concepts, Models and Techniques*. New York: Springer-Verlag
- [14] Ma, J., Xu, W., Sun, Y.H., Turban, E., Wang, S., Liu, O. (2012). *An Ontology-Based Text-Mining Method to Cluster Proposals for Research Project Selection*. IEEE Trans. Syst., Man, Cybern. Part A, Syst & Humans (vol. 42)
- [15] Berry, M. W., dan J. Kogan. (2010). *Text Mining Application and Theory*. United Kingdom: WILEY.
- [16] Hadna, N.M.S., Santosa, P.I., Winarno, W.W. 2016. "Studi literatur tentang perbandingan metode proses analisis sentimen di twitter". Fakultas Teknik, Universitas Gadjah Mada.
- [17] Nugroho, H.T.I. "Pengaruh Algoritma Stemming Nazief-Adriani Terhadap Kinerja Algoritma Winnowing Untuk Mendeteksi Plagiarisme Bahasa Indonesia". ULTIMA Computing. 9 (1):36-40. 2017.
- [18] Baharsyah, I., Cholissodin, I., Setiawan, B.D. "Klasifikasi Deep Sentiment Analysis E-Complain Universitas Brawijaya Menggunakan Metode K-Nearest Neighbor" Universitas Brawijaya Malang, Vol. 3, No. 8, 2014.
- [19] Hartanto, "Text Mining dan Sentimen Analisis Twitter pada Gerakan LGBT" Intuisi Jurnal Psikologi Ilmiah, Intuisi 9 (1) Hal. 18-25, 2017
- [20] Simanjuntak, T.H., Mahmudy, W.F., Sutrisno., "Implementasi Modified K-Nearest Neighbor Dengan Otomatisasi Nilai K Pada Pengklasifikasian Penyakit Tanaman Kedelai" Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Vol. 1, No. 2 Hal. 75-79, Februari 2017
- [21] Hamid, N.A., Nawis, N.M., Ghazali, R., Salleh, M.N.M., "Accelerating Learning Performance of Back Propagation Algorithm by Using Adaptive Gain Together with Adaptive Momentum and Adaptive Learning Rate on Classification Problems". International Journal of Software Engineering and Its Applications, Vol. 5 No. 4. 2011.
- [22] Seminar, K.B., Buono, A., Sukin, T.P.J., "Desain dan Implementasi Komputasi Paralel dengan Algoritma Genetik untuk Prapemrosesan Probabilistic Neural Network" Jurnal Ilmiah Ilmu Komputer, Vol. 3, No. 1, 2005
- [23] Perkasa, T.R., Widyantara, H., Susanto, P., "Rancang Bangun Pendeteksi Gerak Menggunakan Metode Image Subtraction Pada Single Board Computer (SBC)" Journal of Control and Network Systems, Vol. 3 No. 2 Hal. 90-97. 2014.
- [24] Nosrati, M., "Python: An appropriate language for real world programming." World Applied Programming , Vol. 1 No. 2 Hal 110-117. 2011.
- [25] Zulfa, I., Winarko, E., "Sentimen Analisis Tweet Berbahasa Indonesia dengan Deep Belief Network". IJCCS, Vol. 11, No. 2, Hal 187-198, Juli 2017.
- [26] De, S.R., Bandyopadhyay, S.K., "Sentiment Analysis On Product Purchase Through E Commerce". International Journal of Scientific Research and Management (IJSRM), Vol. 5, Hal 5441-5444, Juni 2017.
- [27] Okfalisa Okfalisa, Ikbil Gazalba, Mustakim Mustakim, Nurul Gayatri Indah Reza. "Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification". Information Technology, Information Systems and Electrical Engineering (ICITISEE). Vol 2 No 1. 2017