

Perbandingan Akurasi Naïve Bayes dan K-Nearest Neighbor pada Klasifikasi untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali

M. Azman Maricar¹, Dian Pramana²

Institut Teknologi dan Bisnis (ITB) STIKOM Bali

e-mail: ¹azman@stikom-bali.ac.id, ²dian@stikom-bali.ac.id

Diajukan: 11 September 2019; Direvisi: 25 Oktober 2019; Diterima: 26 Oktober 2019

Abstrak

Tujuan utama setiap alumni setelah lulus di perguruan tinggi adalah mendapatkan pekerjaan. Setiap alumni tentu memiliki rentang waktu yang berbeda untuk mendapatkan pekerjaan setelah lulus, maka dari itu diperlukan metode untuk memprediksi seberapa lama waktu yang diperlukan oleh alumni dalam mendapatkan pekerjaan setelah lulus. Dalam penelitian ini dilakukan perbandingan metode antara Naïve Bayes dan K-Nearest Neighbor dengan nilai $K = 3, 5, 7, \text{ dan } 9$. Parameter yang digunakan sebagai acuan prediksi rentang waktu dalam mendapatkan pekerjaan adalah masa studi, jenis kelamin, dan IPK terakhir. Prediksi yang dihasilkan adalah berupa keterangan cepat yang berarti tiga bulan ke bawah dan lama yang berarti di atas tiga bulan. Data yang digunakan sebanyak 1669, dimana 80% atau 1335 sebagai data training dan 20% atau 334 sebagai data testing. Hasil yang diperoleh adalah Naïve Bayes memiliki akurasi dan MAPE yang lebih baik yaitu 83.83% dan 16.17%, dibandingkan dengan K-Nearest Neighbor dengan nilai K terbaik yaitu 9 yang memiliki akurasi 82.34% dan MAPE 17.66%. Berdasarkan ketentuan rentang nilai MAPE, baik Naïve Bayes dan K-Nearest Neighbor dengan nilai $K=9$ memiliki arti bahwa metode tersebut baik dalam kasus ini, namun Naïve Bayes sedikit lebih baik.

Kata kunci: Peramalan, Naïve Bayes, K-Nearest Neighbor, Akurasi, Mean Absolute Percent Error.

Abstract

The main goal of every graduate after graduating from college is to get a job. Each graduate certainly has a different time span to get a job after graduation, therefore a method is needed to predict how long it will take for graduate to get a job after graduation. In this study a comparison of methods between Naïve Bayes and K-Nearest Neighbor with values $K = 3, 5, 7, \text{ and } 9$. The parameters used as a reference for predicting the time span in getting a job are the study period, gender, and final GPA. The resulting prediction is in the form of a quick statement which means three months down and a long meaning of more than three months. The data used were 1669, of which 80% or 1335 as training data and 20% or 334 as testing data. The results obtained are Naïve Bayes having better accuracy and MAPE, namely 83.83% and 16.17%, compared to K-Nearest Neighbor with the best value of $K = 9$ which has an accuracy of 82.34% and MAPE 17.66%. Based on the MAPE value range, both Naïve Bayes and K-Nearest Neighbor with a value of $K = 9$ means that the method is good in this case, but Naïve Bayes is slightly better.

Keywords: Forecasting, Naïve Bayes, K-Nearest Neighbor, Accuracy, Mean Absolute Percent Error.

1. Pendahuluan

Mahasiswa merupakan aset yang sangat penting bagi institusi pendidikan, maka setiap perguruan tinggi dituntut untuk memberikan pendidikan yang berkualitas agar menghasilkan sumber daya manusia yang memiliki daya saing serta perlu memperhatikan terhadap tingkat kelulusan mahasiswa tepat pada waktunya [1]. Mahasiswa yang telah lulus (alumni) dan perguruan tinggi harus memiliki sinergi yang baik untuk mengetahui eksistensi dari alumninya. Hal ini diperlukan agar setiap perguruan tinggi mampu melakukan perbaikan-perbaikan yang diperlukan untuk menciptakan kualitas dalam segala kegiatan akademik [2].

Institut Teknologi dan Bisnis (ITB) STIKOM Bali merupakan salah satu institusi pendidikan dibidang teknologi informasi yang telah memiliki banyak alumni dan bertambah setiap tahunnya. Keseluruhan alumni tersebut tersebar di berbagai bidang pekerjaan seperti telekomunikasi, *programming*,

hingga wirausaha. Dalam dunia pekerjaan dibutuhkan pengetahuan dan kemampuan yang didapatkan oleh para alumni baik melalui STIKOM Bali maupun melalui pengembangan diri yang dilakukan oleh alumni itu sendiri. Permasalahan yang terjadi adalah seberapa lama para alumni STIKOM Bali mampu mendapatkan pekerjaan.

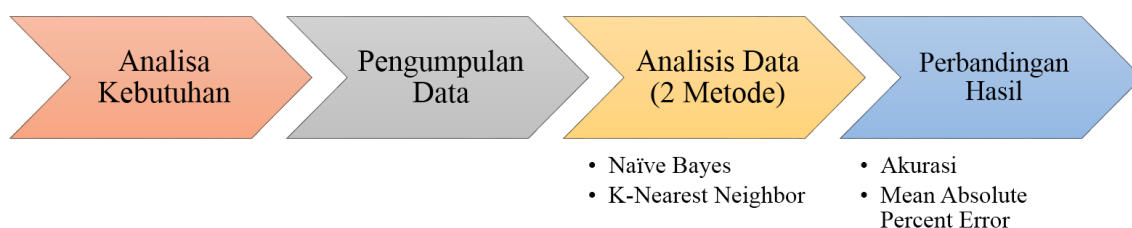
Untuk mengetahui waktu yang diperlukan bagi alumni STIKOM Bali untuk mendapatkan pekerjaan, maka perlu dilakukan peramalan terhadap hal tersebut dengan menggunakan pendekatan *data mining*. Data yang diperlukan adalah data alumni STIKOM Bali yang terdiri dari jenis kelamin, masa studi, IPK terakhir, dan rentang waktu yang diperlukan untuk mendapatkan pekerjaan. Data yang telah terkumpul akan dibagi menjadi dua, yaitu data latih dan data uji. Data uji akan di klasifikasi berdasarkan data latih yang ada untuk mendapatkan rentang waktu yang diperlukan dalam mendapatkan pekerjaan. Proses klasifikasi menggunakan metode Naïve Bayes dan K-Nearest Neighbor. Kedua metode tersebut akan dibandingkan hasilnya berdasarkan akurasi yang didapatkan. Metode dengan akurasi terbaik akan diterapkan dalam sistem peramalan yang dapat digunakan oleh alumni baru lulus dan belum mendapatkan pekerjaan.

Dalam penelitian terdahulu, telah dilakukan beberapa studi terkait kedua metode. Pertama, penggunaan Naïve Bayes dan K-Nearest Neighbor dalam analisis data status kerja di Kabupaten Demak tahun 2012, didapatkan akurasi untuk Naïve Bayes 94,09% dan KNN 96,06% [3]. Kedua, penerapan K-Nearest Neighbor untuk penentuan *grade* sepeda motor, didapatkan akurasi sebesar 64,03% [4]. Ketiga, penggunaan Naïve Bayes untuk prediksi kelulusan, didapatkan hasil 93% [5].

Hasil yang diharapkan melalui penelitian ini adalah mendapatkan akurasi terbaik dari *Naïve Bayes* dan K-Nearest Neighbor, di mana metode dengan akurasi terbaik dapat bermanfaat jika diterapkan dikemudian hari pada suatu sistem peramalan untuk memprediksi status pekerjaan alumni dari segi cepat atau tidaknya alumni tersebut dalam mendapatkan pekerjaan.

2. Metode Penelitian

Beberapa metode atau alur penelitian yang dilakukan dalam penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Metode/alur penelitian.

Berikut adalah penjabaran dari masing-masing alur penelitian yang ditunjukkan pada Gambar 1.

2.1. Analisa Kebutuhan

Pada tahap ini dilakukan analisa terhadap kebutuhan yang terkait dengan penelitian ini, seperti data yang seperti apa yang diperlukan untuk menunjang kegiatan penelitian.

2.2. Pengumpulan Data

Pengumpulan data dilakukan dengan cara mengajukan surat permohonan data hasil *tracer study* alumni, yang diajukan langsung kepada Rektor ITB STIKOM Bali. Data yang digunakan dalam penelitian ini adalah hasil *tracer study* alumni ITB STIKOM Bali, rentang waktu 2017 – 2019, dengan atribut sebagai berikut : nama, angkatan, tahun lulus, masa studi, jenis kelamin, IPK terakhir, dan rentang waktu yang diperlukan untuk mendapatkan pekerjaan. Atribut yang digunakan dalam proses klasifikasi adalah masa studi, jenis kelamin, IPK terakhir, dan rentang waktu yang diperlukan untuk mendapatkan pekerjaan. Total data yang didapatkan 1669.

2.3. Analisis Data (Dua Metode)

Penelitian ini menggunakan pendekatan *data mining*, guna membantu dalam proses klasifikasi/peramalan rentang waktu kerja alumni dengan menggunakan metode *Naïve Bayes* dan *K-Nearest Neighbor* dengan nilai $K = 3, 5, 7$, dan 9 . Kedua metode dibandingkan untuk mendapatkan akurasi terbaik. Aplikasi yang digunakan untuk membandingkan kedua metode adalah WEKA.

2.3.1. Data Mining

Dikenal sebagai *Knowledge Discovery in Database (KDD)*, *data Mining* merupakan sebuah proses untuk menemukan hubungan atau pola yang memiliki makna dengan cara mengamati data yang besar, di mana datanya disimpan dalam sebuah penyimpanan, dan mengelolanya dengan teknik statistik dan/atau matematika guna mendapatkan manfaat [6]. *Data mining* pada saat ini dapat mengatasi jumlah data sangat besar/banyak, dimensi data yang sangat tinggi dan kumpulan data yang beragam [6].

Metode-metode dalam pengolahan *data mining* [7]:

1. *Predictive modelling* merupakan pengolahan *data mining* dengan melakukan prediksi. Tujuannya untuk membangun model prediksi suatu nilai dengan ciri-ciri tertentu. Contoh algoritmanya *Linear Regression*, *Neural Network*, *Support Vector Machine*, dan lain-lain.
2. *Association* (Asosiasi) merupakan teknik *data mining* yang mempelajari hubungan antar data. Contoh penggunaannya seperti untuk menganalisis perilaku mahasiswa yang datang terlambat. Contoh algoritmanya *FP-Growth*, *A Priori*, dan lain-lain.
3. *Clustering* (Klastering) atau pengelompokan merupakan teknik untuk mengidentifikasi data ke dalam suatu kelompok tertentu. Contoh algoritmanya *K-Means*, *K-Medoids*, *Self-Organisation Map (SOM)*, *Fuzzy C-Means*, dan lain-lain.
4. *Classification* merupakan teknik mengklasifikasikan data. Perbedaan mendasar dengan *clustering* adalah data, jika pada *clustering* tidak ada variabel dependen, sedangkan *classification variable dependent* harus ada. Contoh algoritma yang menggunakan metode ini *ID3* dan *K Nearest Neighbors*

2.3.2. Peramalan

Peramalan merupakan suatu ilmu untuk memprediksi sesuatu yang belum terjadi atau memprediksi sesuatu yang akan terjadi di masa depan. Peramalan sangat penting dan dapat diterapkan dalam berbagai bidang [6].

Prosedur-prosedur dalam peramalan, yaitu [6]:

1. Analisis Masalah
2. Pengumpulan Data
3. Analisis Data
4. Pemilihan Metode
5. Penerapan Metode
6. Pengujian Metode
7. *Monitoring Kinerja Metode*

Metode umum dalam peramalan ada dua yaitu kualitatif dan kuantitatif, di mana kedua metode tersebut memiliki sifat masing-masing. Kualitatif sifatnya intuitif dan tidak memiliki data masa lalu, sehingga tidak dapat diselesaikan dengan cara matematika, karena lebih banyak menggunakan pendapat tertentu. Hal tersebut berbeda dengan kuantitatif yang memiliki data *history*, dan dapat dihitung secara matematis. Dalam penggunaannya metode kuantitatif lebih sering digunakan daripada kualitatif, yang lebih dikenal sebagai *time series*, yang bentuk datanya terakumulasi selama periode waktu tertentu. *Time series* memiliki empat komponen penting yaitu, Tren Sekuler, Variasi Musim, Variasi Siklus, dan Variasi Ireguler. Penggunaan *time series*, cenderung digunakan untuk memprediksi masa depan serta dibuat menggunakan *set data* yang rinci yang dihasilkan dari di masa lalu [6].

2.3.3. Naïve bayes

Naïve Bayes adalah metode klasifikasi atau teknik *machine learning* yang populer/umum digunakan dalam klasifikasi teks, memiliki performa yang baik pada banyak domain, sederhana dan efisien. Namun *Naïve Bayes* sangat sensitif dalam pemilihan fitur, terlalu banyak jumlah fitur dapat mengakibatkan meningkatnya waktu perhitungan dan menurunkan akurasi klasifikasi [8].

Langkah-langkah dalam algoritma *Naïve Bayes* [8]:

1. Hitung probabilitas bersyarat/*likelihood*

$$P(x \rightarrow | C_i) = x_1, x_2, \dots, x_n | C \quad (1)$$

Dimana :

$C = class$

$x =$ vektor dari nilai atribut n

$P(x_i | C) =$ proporsi dokumen dari *class* C yang mengandung nilai atribut x_i

2. Hitung probabilitas *prior* untuk tiap kelas

$$P(C_i) = \frac{N_j}{N} \tag{2}$$

Dimana :

N_j : Jumlah dokumen dalam suatu *class*

N : Jumlah total dokumen

3. Hitung probabilitas *posterior*

$$Posterior = \frac{likelihood \times prior}{evidence} \tag{3}$$

2.3.4. K-Nearest Neighbor

K-Nearest Neighbor merupakan metode klasifikasi yang berdasarkan atribut dan *training sample*. Nilai objek yang akan diklasifikasi akan dicari jarak terdekatnya berdasarkan nilai k yang ditentukan, klasifikasinya menggunakan *voting* terbanyak dari nilai k yang ditentukan, dengan arti lain metode *K-Nearest Neighbor* merupakan metode klasifikasi yang menggunakan ketetanggaan sebagai nilai prediksi terhadap nilai objek [9].

2.4. Perbandingan Hasil

Perbandingan hasil terhadap kedua metode dilakukan dengan metode pengukuran hasil klasifikasi/peramalan yaitu, akurasi dan *mean absolute Percent Error* (MAPE).

2.4.1. Akurasi

Pengujian akurasi merupakan pengujian yang dilakukan untuk memperkirakan seberapa tepat hasil klasifikasi terhadap data yang ada, di mana persamaan 4 digunakan untuk menghitung akurasi [10].

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{4}$$

Dimana :

TP = Hal positif yang diklasifikasikan dengan benar

TN = Hal negatif yang diklasifikasikan dengan benar

FP = Hal positif yang diklasifikasikan dengan salah

FN = Hal negatif yang diklasifikasikan dengan salah

2.4.2. Mean Absolute Percent Error (MAPE)

Mean Absolute Percent Error (MAPE) merupakan perhitungan yang dilakukan untuk mengetahui rata-rata kesalahan absolut dalam bentuk persentase, di mana perhitungannya dapat dilakukan dengan persamaan 5 [6]:

$$MAPE = \sum \frac{|A-F|}{n} \times 100\% \tag{5}$$

Dimana:

A : Nilai Aktual

F : Nilai Peramalan

n : Jumlah Periode yang digunakan

MAPE memiliki standar penilaian terhadap hasil tingkat *error* yang dihasilkan oleh perhitungannya [6], hal ini ditunjukkan pada Tabel 1.

Tabel 1. Rentang nilai MAPE.

Rentang Nilai	Pengertian
<10%	Kemampuan metode sangat baik
10-20%	Kemampuan metode baik
20-50%	Kemampuan metode buruk
>50%	Kemampuan metode sangat buruk

3. Hasil dan Pembahasan

Sebelum melakukan perhitungan akurasi dan MAPE dari kedua metode, hal yang pertama dilakukan adalah melakukan perbandingan terhadap perubahan data pada salah satu atribut dan menentukan perbandingan yang optimal antara data *training* dan *testing*.

3.1. Perubahan dan Perbandingan Data

Dari atribut data yang didapatkan (nama, angkatan, tahun lulus, masa studi, jenis kelamin, IPK terakhir, dan rentang waktu yang diperlukan untuk mendapatkan pekerjaan), hanya masa studi, jenis kelamin, IPK terakhir, dan rentang waktu yang digunakan sebagai parameter untuk klasifikasi dengan menggunakan aplikasi Weka dan menggunakan metode *Naïve Bayes* dan *K-Nearest Neighbor*, dimana nilai K pada *K-Nearest Neighbor* yang diujikan adalah 3, 5, 7, dan 9. Data dari masing-masing atribut yang digunakan ditunjukkan pada Tabel 2.

Tabel 2. Data masing-masing atribut.

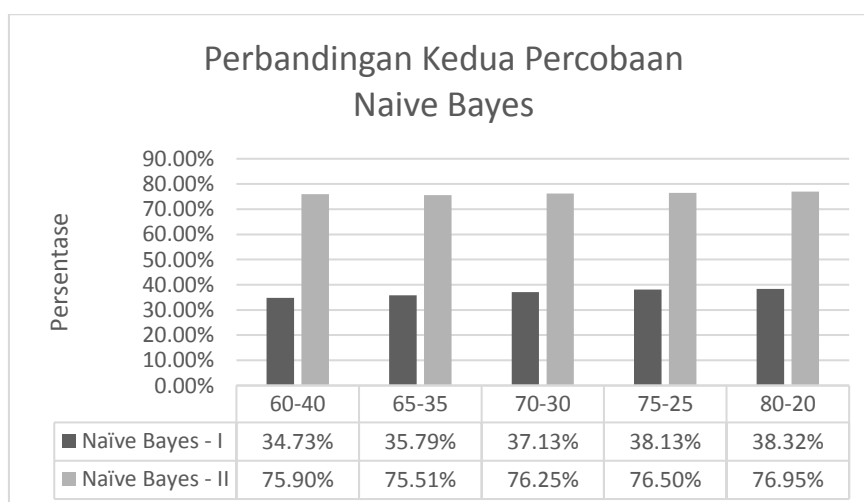
Atribut	Data
Masa Studi	1.5 – 8.5 (Tahun)
Jenis Kelamin	Laki-Laki & Perempuan
IPK Terakhir	2.55 – 3.94
Rentang Waktu	0 Bulan, 1 Bulan, 2 Bulan, 3 Bulan, Di Atas 3 Bulan

Perubahan data dilakukan terhadap data yang ada pada atribut Rentang Waktu. Perubahan tersebut ditunjukkan pada Tabel 3.

Tabel 3. Perubahan data atribut rentang waktu.

Data Asli	Perubahan
0 Bulan	Cepat
1 Bulan	
2 Bulan	
3 Bulan	
Di atas 3 Bulan	Lama

Percobaan pertama dilakukan dengan menggunakan data seperti pada tabel 2 dan percobaan kedua dilakukan dengan perubahan seperti yang ditunjukkan pada Tabel 3. Terjadi peningkatan yang signifikan dari percobaan pertama ke percobaan kedua. Pada percobaan pertama akurasi terbesar adalah 38.32%, sedangkan pada percobaan kedua akurasi terbesar adalah 76.95%, di mana pada kedua percobaan tersebut, hasil yang lebih optimal adalah perbandingan 80% data *training* dan 20% data *testing*. Berdasarkan peningkatan akurasi yang signifikan, yang dipengaruhi perubahan data yang dilakukan pada atribut rentang waktu, dapat dikatakan kedua metode tidak optimal terhadap jumlah *class* yang terlalu banyak. Untuk menunjukkan peningkatan hasil yang signifikan pada kedua percobaan, sebagai *sample* ditampilkan peningkatan hasil tersebut untuk metode *Naïve Bayes* pada Gambar 2.



Gambar 2. Peningkatan akurasi Naïve Bayes.

3.2. Perhitungan Akurasi dan MAPE

Berdasarkan hasil pada perubahan dan perbandingan data, maka model data yang digunakan adalah seperti yang dilakukan pada percobaan kedua dengan perbandingan data 80% data *training* dan 20% data *testing*. Sehingga total data *training* yang digunakan adalah 1335 dan data *testing* sebesar 334 dari total 1669 data.

Dengan menggunakan aplikasi WEKA maka hasil klasifikasi/prediksi masing-masing metode ditunjukkan pada Tabel 4. Di mana akurasi dihitung dengan menggunakan persamaan 4 dan MAPE dengan persamaan 5.

Tabel 4. Hasil klasifikasi/prediksi.

Metode	Terklasifikasi Benar	Terklasifikasi Salah	Akurasi	MAPE
<i>Naïve Bayes</i>	280	54	$Akurasi = \frac{280}{334} \times 100\%$ Akurasi = 83.83%	$MAPE = \frac{54}{334} \times 100\%$ MAPE = 16.17%
<i>K-Nearest Neighbor 3</i>	254	80	$Akurasi = \frac{254}{334} \times 100\%$ Akurasi = 76.05%	$MAPE = \frac{80}{334} \times 100\%$ MAPE = 23.95%
<i>K-Nearest Neighbor 5</i>	263	71	$Akurasi = \frac{263}{334} \times 100\%$ Akurasi = 78.74%	$MAPE = \frac{71}{334} \times 100\%$ MAPE = 21.26%
<i>K-Nearest Neighbor 7</i>	272	62	$Akurasi = \frac{272}{334} \times 100\%$ Akurasi = 81.44%	$MAPE = \frac{62}{334} \times 100\%$ MAPE = 18.56%
<i>K-Nearest Neighbor 9</i>	275	59	$Akurasi = \frac{275}{334} \times 100\%$ Akurasi = 82.34%	$MAPE = \frac{59}{334} \times 100\%$ MAPE = 17.66%

Berdasarkan tabel 4, maka akurasi terbaik adalah metode *Naïve Bayes* dengan akurasi 83.83% dan MAPE sebesar 16.17%. Berbeda tipis dengan metode *K-Nearest Neighbor* dengan nilai $K = 9$ yaitu, 82.34% dengan MAPE 17.66%. Merujuk pada tabel 1 mengenai rentang Nilai MAPE, persentase 16.17% berada direntang nilai 10-20%, yang berarti kemampuan metode *Naïve Bayes* pada kasus ini adalah baik.

Dengan menggunakan tiga parameter yaitu, jenis kelamin, masa studi, dan IPK terakhir, baik metode *Naïve Bayes* maupun *K-Nearest Neighbor* dengan nilai $K = 3, 5, 7, \text{ dan } 9$, mampu menjadi solusi dari permasalahan yang ada yaitu untuk mengetahui waktu yang diperlukan bagi alumni STIKOM Bali untuk mendapatkan pekerjaan dengan cara melakukan klasifikasi/peramalan terhadap kasus tersebut. Namun dari hasil pengukuran akurasi dan MAPE, metode *Naïve Bayes* memiliki hasil yang lebih baik dibandingkan dengan metode *K-Nearest Neighbor*.

4. Kesimpulan

Berdasarkan pada percobaan-percobaan yang telah dilakukan serta perhitungan akurasi dan MAPE terhadap kedua metode tersebut, diperoleh hasil sebagai berikut :

1. Kedua metode, baik metode *Naïve Bayes* maupun *K-Nearest Neighbor* dengan nilai $K = 3, 5, 7, \text{ dan } 9$, mampu menjadi solusi dari permasalahan yang ada yaitu untuk mengetahui waktu yang diperlukan bagi alumni STIKOM Bali untuk mendapatkan pekerjaan dengan cara melakukan klasifikasi/peramalan terhadap kasus penelitian ini, di mana parameter yang digunakan adalah jenis kelamin, masa studi, dan IPK terakhir.
2. Terjadi peningkatan akurasi yang signifikan terhadap kedua metode dari percobaan pertama ke percobaan kedua. Peningkatan tersebut dipengaruhi oleh perubahan data pada atribut rentang waktu, di mana pada percobaan pertama jumlah *class* pada atribut tersebut adalah lima, sedangkan pada percobaan kedua hanya memiliki 2 *class*. Akibat dari peningkatan tersebut, dapat dikatakan bahwa kedua metode tidak maksimal pada jumlah *class* yang banyak.
3. Dengan menggunakan 1335 data *training* dan 334 data *testing*, *Naïve Bayes* memiliki akurasi dan MAPE yang lebih baik yaitu 83.83% dan 16.17%, dibandingkan dengan *K-Nearest Neighbor* dengan nilai K terbaik yaitu 9 yang memiliki akurasi 82.34% dan MAPE 17.66%.
4. Berdasarkan ketentuan rentang nilai MAPE, baik *Naïve Bayes* dan *K-Nearest Neighbor* dengan nilai $K=9$ memiliki arti bahwa metode tersebut baik dalam kasus ini, namun *Naïve Bayes* sedikit lebih baik.

Daftar Pustaka

- [1] A. Y. Saputra and Y. Primadasa, "Penerapan Teknik Klasifikasi untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *Techno.COM*, Vol. 17, No. 4, pp. 395-403, 2018.
- [2] Z. A. Nugroho and R. Arifudin, "Sistem Informasi Tracer Study Alumni Universitas Negerri Semarang dengan Aplikasi Digital Maps," *Scientific Journal of Informatics Vol. 1, No. 2*, pp. 153-160, 2014.
- [3] R. E. Putri, Suparti and R. Rahmawati, "Perbandingan Metode Klasifikasi Naïve Bayes dan K-Nearest Neighbor Pada Analisis Data Status Kerja Di Kabupaten Demak Tahun 2012," *JURNAL GAUSSIAN*, Vol. 3, No. 4, pp. 831 - 838, 2014.
- [4] H. Leidiyana, "Penerapan Metode K-Nearest Neighbor Pada Penentuan Grade Dealer Sepeda Motor," *Jurnal Ilmu Pengetahuan dan Teknik Komputer*, Vol.2, No.2, pp. 108-112, 2017.
- [5] D. L. Fithri and E. Darmanto, "Sistem Pendukung Keputusan Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Metode Naïve Bayes," in *Prosiding SNATIF Ke-1*, 2014.
- [6] M. A. Maricar, P. Widiadnyana and W. A. Wijaya, "Analysis of Data Mining for Forecasting Total Goods Delivery with Moving Average Method (Case Study: Agent "X" Expedition "Z")," *International Journal of Engineering and Emerging Technology*, Vol. 2, No. 1, pp. 7-10, 2017.
- [7] D. H. Kamagi and S. Hansun, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *ULTIMATICS*, Vol. VI, No. 1, pp. 15-20, 2014.
- [8] D. A. Muthia, "Opinion Mining pada Review Buku Menggunakan Algoritma Naive Bayes," *Jurnal Teknik Komputer Amik BSI*, Vol.II, No.1, pp. 1-8, 2016.
- [9] S. K. Lidya, O. S. Sitompul and S. Efendi, "Sentiment Analysis pada Teks Bahasa Indonesia Menggunakan Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN)," in *SENTIKA*, Yogyakarta, 2015.
- [10] M. A. Maricar, I. N. S. Kumara and M. Sudarma, "Opinion Mining on Twitter Social Media to Classify Racism Using Combination of POS Tagging, Naive Bayes Classifier, and K-Nearest Neighbor," in *International Conference on Smart-Green Technology in Electrical and Information System*, Bali, 2018.