

# PENERAPAN ALGORITMA *K-MEANS* DAN ALGORITMA *NAÏVE BAYES* DALAM PEMILIHAN KONSENTRASI JURUSAN SISWA DI SMA ISLAM TA'ALUMUL HUDA BUMIAYU

M Ghozi Ghozali<sup>1</sup>, Nurul Mega Saraswati<sup>2</sup>, Rito Cipta Sigitta Hariyono<sup>3</sup>

<sup>1</sup>Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Peradaban,

<sup>2</sup>Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Peradaban,

<sup>3</sup>Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Peradaban.

Ghozighozali45@gmail.com, nurulmega@peradaban.ac.id, ritocipta@peradaban.ac.id

Jl. Raya Pagojengan KM 03 Paguyangan Brebes

## **Keyword**

*K-Means, Naïve Bayes, concentration majors, prediction, Confusion Matrix*

## **Abstract**

*The determination of majors in high school, vocational high school and Madrasah Aliyah is done at the beginning of school. Sma Islam Ta'alumul Huda Bumiayu implemented the 2013 education unit level curriculum as a foundation in the cornering process. The selection of the concentration of majors should take into account the interests and academic ability of students so as not to have difficulty in absorbing and understanding the material, to overcome the above problems is required methods that can provide reference for students in selecting the concentration of majors. The methods used are the K-Means clustering algorithm and the Naïve Bayes algorithm. Classification results using the K-Means clustering algorithm and Naive Bayes obtained 80% accuracy.*

## **Kata Kunci**

*K-Means, Naïve Bayes, konsentrasi jurusan, prediksi, Confusion Matrix*

## **Abstrak**

Penentuan jurusan pada sekolah menengah atas, sekolah menengah kejuruan maupun madrasah Aliyah dilakukan pada awal masuk sekolah. SMA Islam Ta'alumul Huda Bumiayu menerapkan kurikulum tingkat satuan pendidikan 2013 sebagai landasan dalam proses penjurusan. Pemilihan konsentrasi jurusan harus mempertimbangkan minat dan kemampuan akademik siswa agar tidak kesulitan dalam menyerap dan memahami materi, untuk mengatasi masalah diatas diperlukan metode yang dapat memberikan acuan bagi siswa dalam memilih konsentrasi jurusan. Metode yang digunakan adalah algoritma *K-Means clustering* dan algoritma *Naïve Bayes*. Hasil klasifikasi menggunakan algoritma *K-Means clustering* dan *Naive Bayes* memperoleh akurasi sebesar 80%.

## **PENDAHULUAN**

Perkembangan pada era 4.0 membawa perubahan yang begitu pesat dan berkembang hampir mencakup berbagai sektor. Perkembangan teknologi menghasilkan data yang besar dari berbagai bidang, baik dalam bidang industri, bidang ekonomi, bidang kesehatan, dan bidang lainnya, termasuk penerapan teknologi dalam bidang pendidikan yang dapat menghasilkan data yang cukup besar. Seiring dengan perkembangan teknologi, banyak sekali data yang dapat dimanfaatkan dan diolah menjadi sebuah informasi, termasuk teknologi informasi yang dapat membantu siswa dalam menentukan penjurusan pada tingkat Sekolah Menengah Atas (SMA) atau sederajat. Saat ini sebagian besar siswa Sekolah Menengah Pertama (SMP) yang akan melanjutkan pendidikan ke jenjang selanjutnya yakni Sekolah Menengah Atas (SMA) cenderung belum mengetahui minat dan kemampuannya, sehingga siswa merasa kesulitan dalam memilih konsentrasi jurusan yang akan di pilih pada saat memasuki SMA. Pemilihan konsentrasi jurusan sangat penting karena mempengaruhi proses belajar siswa ke depannya, sehingga dalam memilih konsentrasi harus dipertimbangkan secara matang sesuai minat dan kemampuan akademik siswa. Beberapa kasus yang sering terjadi siswa memilih jurusan yang kurang tepat, sehingga menyebabkan siswa kesulitan dalam mengikuti dan memahami materi-materi yang

diberikan, hal ini akan mempengaruhi proses belajar siswa yang tidak maksimal dan menurunnya prestasi akademik siswa karena siswa merasa kurang tepat dengan jurusan yang di pilihnya. SMA Islam Ta'alumul Huda Bumiayu. Sekolah ini terletak di kabupaten Brebes bagian selatan beralamatkan di Jl. KH. Ahmad Dahlan no. 99 Desa Bumiayu Kecamatan Bumiayu Kabupaten Brebes. SMA Islam Ta'alumul Huda Bumiayu memiliki tiga (3) jurusan yaitu jurusan IPA, jurusan IPS dan jurusan Bahasa. Sistem penentuan jurusan sekolah ini merujuk pada peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 81A tahun 2013 tentang Implementasi kurikulum mengemukakan bahwa penentuan jurusan pada sekolah menengah atas, sekolah menengah kejuruan maupun madrasah Aliyah di lakukan langsung pada awal masuk sekolah atau pada saat siswa kelas sepuluh (X)[1]. SMA Islam Ta'alumul Huda Bumiayu merupakan salah satu SMA yang siswa siswinya cenderung banyak yang ragu dalam memilih konsentrasi jurusan yang akan di pilih. Observasi dengan pihak SMA Islam Ta'alumul Huda mendapat informasi, SMA tersebut tidak menjadikan nilai ujian nasional sebagai acuan untuk menentukan jurusan bagi siswa. Sebelum menentukan jurusan, siswa/siswi wajib mengikuti *pretest* tertulis sebagai acuan untuk menentukan jurusannya. Hal ini membantu siswa yang masih ragu dalam menentukan jurusan mana yang akan di pilih, sehingga untuk meminimalisir memilih jurusan hanya mengikuti teman ataupun dorongan dari orang tua yang nantinya mengakibatkan siswa merasa salah memilih jurusan. Pada bagian kurikulum SMA Islam Ta'alumul Huda, masih kesulitan dalam mengelompokkan siswanya karena hasil *pretest* tidak sesuai dengan kriteria dari jurusan yang diminati siswanya.

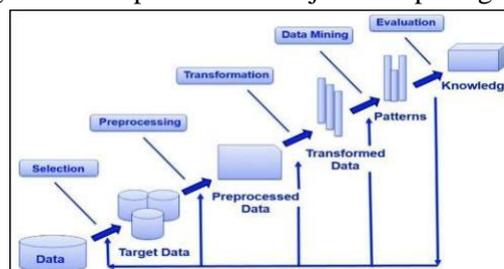
## TINJAUAN PUSTAKA

### A. Kelas Konsentrasi

Konsentrasi menurut KBBI (2014) merupakan pemusatan perhatian atau pikiran terhadap suatu hal, sedangkan kelas konsentrasi adalah suatu penjurusan yang menitik beratkan pada satu di bidang tertentu [8].

### B. Data Mining

Data *mining* merupakan bidang dari beberapa keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, *database*, dan visualisasi sebagai penanganan permasalahan pengambilan informasi dari database yang besar[9]. Hasil yang didapat dari proses data *mining* dapat diterapkan diberbagai sektor seperti bisnis, pendidikan, kesehatan dan lainnya. Proses menggali informasi dari data *mining* melibatkan teknik dari beberapa disiplin ilmu seperti teknologi *database* dan data *warehouse*, statistik, *machine learning*, komputasi dengan kinerja tinggi, *pattern recognition*, *neural network*, visualisasi data.[10] *Data mining* ialah bagian dari KDD (*Knowledge Discovery in Database*)[11]. Langkah -langkah dalam proses KDD dijelaskan pada gambar 2.1:



Gambar 1 Tahapan KDD dalam *Data Mining*

Adapun proses pada *knowledge data discovery* (KDD) secara garis besar dijelaskan sebagai berikut ;

#### 1. *Data selection*

Seleksi data dari sekumpulan data operasional sebelum tahap penggalian informasi dalam KDD dimulai.

#### 2. *Pre-processing/cleaning*

Sebelum proses *Data Mining* dilaksanakan Proses pembersihan pada data yang menjadi fokus KDD dengan tujuan untuk membersihkan data dari duplikasi, inkonsisten data, dan tipografi (kesalahan cetak).

#### 3. *Transformation*

*Coding* adalah transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *Data Mining*.

#### 4. *Data Mining*

Proses pencarian pola atau informasi pada suatu data menggunakan teknik atau metode tertentu, pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. Interpretation / Evaluation

Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelum nya.

**C. K-Means**

*K-Means* adalah metode *clustering* atau pengelompokkan data *non-hieraki* yang mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. *K-Means* mempartisi data ke dalam kelompok yang mempunyai karakteristik sama. Adapun tujuan pengelompokkan data ini adalah untuk meminimalkan fungsi objektif yang di *set* dalam suatu kelompok dan memaksimalkan variasi antar kelompok.[5]. Langkah langkah dalam proses algoritma *K-Means* sebagai berikut:

1. Memilih nilai *centroid* secara *random* sebagai pusat *cluster*.
2. Jarak antara data dengan pusat *cluster* dihitung dengan *Euclidian Distance*

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

3. Data dikelompokkan sesuai dengan jarak terdekat dari pusat cluster.
4. Pusat cluster baru akan ditentukan bila semua data telah ditetapkan dalam cluster terdekat.
5. Ulangi proses ke empat sampai nilai centroid tidak berubah lagi.

**D. Naïve Bayes**

Algoritma klasifikasi *Naïve Bayes* (NB) adalah pengklasifikasi yang berdasarkan probabilitas bersyarat pada *teorema Bayes*. NB dapat mengolah data dalam jumlah besar untuk memprediksi probabilitas di masa depan. NB memiliki kelebihan seperti mudah digunakan, satu kali *scan* data *training*, penanganan nilai *atribut* yang hilang, dan data kontinu.[12]

**E. Android**

Android merupakan sistem operasi pada handphone, bersifat terbuka dan berbasis pada sistem operasi linux. Pengembang yang ingin menciptakan aplikasi, android menyediakan platform terbuka [13].

**F. MySQL**

MySQL adalah perangkat lunak sistem manajemen basis data Structured Query Language (SQL), multi-user yang telah banyak terpasang di seluruh dunia. MySQL merupakan dua bentuk lisensi, yaitu FreeSoftware dan Shareware. MySQL merupakan sebuah database server yang free dibawah Lisensi GNU/GPL (General Public License). Database MySQL merupakan suatu perangkat lunak database yang berbentuk database relasional atau disebut Relational Database Management System (RDBMS) yang 5 M Ghazi Gozali / IJIR Vol. 1 No. 1. 2021 menggunakan bahasa permintaan yang bernama SQL (Structured Query Language)[15].

**G. Xampp**

Xampp merupakan software web server apache yang didalamnya berisi database server MySQL dan mendukung pemrograman PHP. Xampp berfungsi sebagai server yang berdiri sendiri (localhost), yang terdiri atas program Apache HTTP server, MySQL database dan penerjemah Bahasa yang ditulis dengan Bahasa pemrograman PHP dan Perl. Xampp dapat diunduh secara gratis dan mendukung instalasi pada OS Linux dan Windows[16].

**H. Hypertext Preprocessor (PHP)**

PHP atau Hypertext Preprocessor ialah sebuah bahasa pemrograman open source untuk membuat dan mengembangkan situs web. Bahasa pemrograman ini dapat digunakan bersamaan dengan Bahasa pemrograman HTML atau Hypertext Markup Language sekaligus bekerja disisi server (server side HTML-embedded scripting). PHP dirancang untuk dapat bekerjasama dengan database server dan pembuatan dokumen HTML yang dapat mengakses database dengan mudah. PHP tidak terbatas pada hasil keluaran HTML, PHP juga dapat mengolah keluaran gambar, file PDF, dan movies Flash[16].

**I. CodeIgniter**

CodeIgniter merupakan sebuah aplikasi pengembangan framework atau sebuah tools yang diperuntukkan bagi programmer yang akan membangun aplikasi berbasis web dengan Bahasa pemrograman PHP. Dengan menggunakan framework, pengembangan proyek dapat dilakukan dengan lebih cepat dan teratur karena disertai dengan konsep Model View Control (MVC). MVC ini membagi pengembangan aplikasi berdasarkan komponen utamanya. Komponen tersebut yaitu manipulasi data

(model), tampilan antar muka (view) dan bagian kontrol dari aplikasi (controller). Framework CodeIgniter juga menyediakan tampilan antar muka yang sederhana[17]

**J. Confusion Matrix**

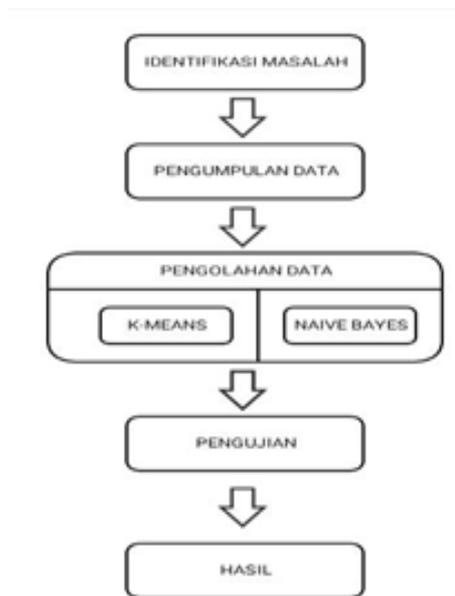
Confusion Matrix adalah alat untuk menganalisis seberapa baik classifier mengenali tuple dari kelas yang berbeda[19]. True Positive dan True Negative memberikan informasi ketika classifier benar, sedangkan False Positive dan False Negative memberikan informasi ketika classifier salah

**K. Root mean Square Error**

Root Mean Square Error (RMSE) adalah alat seleksi model berdasarkan pada error hasil estimasi[20]. RMSE merupakan cara yang populer didalam mengevaluasi hasil prediksi, error yang ada menjelaskan seberapa besar perbedaan hasil estimasi dengan hasil yang akan di estimasi. Nilai RMSE rendah dapat dikatakan akurat, karena nilai yang dihasilkan dari suatu model prediksi mendekati nilai observasinya.

**METODOLOGI**

Metode dari penelitian ini adalah metode eksperimen, dan adapun tahapan-tahapan dari penelitian ini dapat dilihat pada gambar



Gambar 2. Tahapan Penelitian

**1. Identifikasi Masalah**

Pada tahap ini, mengidentifikasi permasalahan yang muncul dan ditempuh dengan melakukan kegiatan survei secara langsung pada SMA Islam Ta'allumul Huda Bumiayu, survei tersebut di laksanakan untuk mengetahui permasalahan yang ada, dari kegiatan survei tersebut menghasilkan sebuah informasi bahwa hasil dari belajar mereka selama di sekolah termasuk nilai akademik masih belum bisa untuk menentukan pilihan mereka didalam memilih jurusan pada jenjang perkuliahan.

**2. Pengumpulan data**

Penelitian diawali dengan melakukan pengumpulan data. Data yang diperoleh adalah data nilai *pretest* siswa-siswi di SMA Ta'alumul Huda Bumiayu tahun ajaran 2019/2020 sebanyak 148 siswa.

**3. Pengolah data awal**

Setelah data terkumpul, kemudian data dinormalisasi. Dari *dataset* diatas akan dibagi menjadi data *sampel* 133 dan untuk data uji 15. Menyeleksi *atribut* data yang akan digunakan dan data yang tidak relevan dihilangkan.

**4. Model/ metode yang diusulkan**

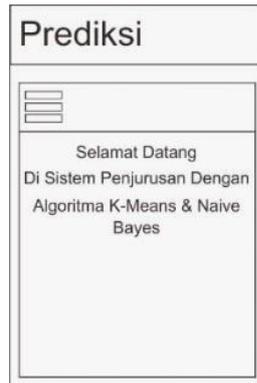
Model dalam penelitian ini berupa metode dalam teknik prediksi *data mining* yaitu pengelompokkan dan klasifikasi dengan algoritma *K-Means* dan *Naïve Bayes* yang dioptimalkan untuk prediksi.

**HASIL DAN PEMBAHASAN**

Proses penggalian informasi atau pengolahan terhadap data yang telah terkumpul dari semua nilai siswa selanjutnya data di olah dengan menerapkan KDD (*knowledge data discovery*) untuk mencari pola maupun informasi dari suatu data tersebut, Berdasarkan penjabaran pada bab 2 tentang tahapan KDD dalam Data Mining maka langkah-langkah pada tahapan ini adalah *Data Selection, Data Cleaning, Transformasion*, proses *Data Mining* yang terdiri dari penerapan algoritma K-Modes dan Algoritma *Naïve Bayes, Interpretation* serta Pengujian. Hasil dari penelitian ini adalah aplikasi yang dapat memprediksi jurusan kuliah siswa-siswi SMA Islam Ta'allumul Huda Bumiayu berdasarkan NB dalam pemilihan konsentrasi jurusan siswa di SMA Islam Ta'alumul Huda Bumiayu, desain tampilan dari program aplikasi sebagai berikut.

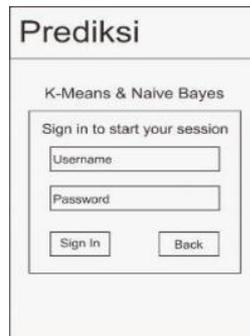
**Desain Tampilan**

**Desain Rancangan Tampilan Awal**



Gambar 3. Tampilan Awal

**Desain Rancangan Halaman Login**



Gambar 4. Halaman Login

**Desain Rancangan Halaman Utama**



Gambar 5. Halaman Utama

### Desain Rancangan Halaman *Input Data*

Gambar 6. Input data

### Desain Rancangan Halaman Hasil

No	Nama**	MTK**	IPA**
1	Niya	4	2
2	Tun	1	5
3	Amalyah	4	1

Gambar 7. Halaman Hasil

## KESIMPULAN DAN SARAN

### Kesimpulan

Hasil dari penelitian yang telah dilakukan oleh penulis, maka dapat disimpulkan bahwa konsentrasi jurusan siswa-siswi SMA Islam Ta'alumul Huda Bumiayu dapat di prediksi menggunakan teknik data mining menggunakan algoritma K-Means dan algoritma Naïve Bayes . Pengolahan data yang dilakukan pada 148 record data dari masingmasing nilai data pretest siswa-siswa SMA Islam Ta'alumul Huda Bumiayu tahun ajaran 2018/2019 menggunakan algoritma K-means dan algoritma Naïve Bayes di dapatkan akurasi sebesar 80% dengan predikat good classification.

### Saran

Saran yang dapat disampaikan kepada pihak-pihak yang berkepentingan dalam mengembangkan dan menyempurnakan lebih lanjut mengenai penerapan algoritma K-Means dan algoritma Naïve Bayes dalam pemilihan konsentrasi jurusan siswa di SMA Islam Ta'alumul Huda Bumiayu adalah sebagai berikut:

1. Pengembangan penelitian ini dengan menambahkan metode optimasi untuk meningkatkan nilai akurasi dalam pemilihan konsentrasi jurusan siswa di SMA Islam Ta'alumul Huda Bumiayu.
2. Mengembangkan aplikasi agar bias berjalan pada lintas platform salah satunya IOS (Apple).

### Daftar Pustaka

- [1] Sabariah, M. Istoningtyas, and Sharipuddin, "Penentuan Jurusan ke Perguruan Tinggi Menggunakan Metode Clustering di SMAN 3 Kuala Tungkal," vol. 13, no. 2, 2018.
- [2] Y. Syahra, "Penerapan Data Mining Dalam Pengelompokkan Data Nilai Siswa Untuk Penentuan Jurusan Siswa Pada SMA Tamora Menggunakan Algoritma K-Means Clustering," vol. 17, no. SAINTIKOM, pp. 228– 233, 2018.
- [3] A. Saleh and F. Nasari, "Penggunaan Teknik Unsupervised Discretization pada Metode Naive Bayes dalam Menentukan Jurusan Siswa Madrasah Aliyah," J. Teknol. Inf. dan Ilmu Komput., vol. 5, no. 3, p. 353, 2018.
- [4] D. Agustini, M. Farida, and A. Rahman, "(1) (2) (3)," vol. 03, no. 02, pp. 58–61, 2018.

- [5] Widodo and D. Wahyuni, "Implementasi algoritma k-means clustering untuk mengetahui bidang skripsi mahasiswa multimedia pendidikan teknik informatika dan komputer universitas negeri jakarta," *J. Pint.*, vol. 01, no. September, p. 11 pages, 2018.
- [6] A. L. R. Khadafy, "Penerapan Naive Bayes untuk Mengurangi Data Noise pada Klasifikasi Multi Kelas dengan Decision Tree," *J. Intell. Syst.*, vol. 1, no. 2, pp. 136–142, 2015.
- [7] P. Seminar, N. Riset, E. Buulolo, and R. Syahputra, "Implementasi Algoritma Clustering K-Means Untuk Mengelompokkan Mahasiswa Baru Yang Berpotensi ( Studi Kasus : Stmik Budi Darma )," vol. 2, no. September, pp. 17–24, 2019.
- [8] N. Rusmania, "PERANCANGAN APLIKASI K-MEANS SEBAGAI PENENTU KONSENTRASI BAGI MAHASISWA INFORMATIKA UMS," vol. 151, pp. 10–17, 2015.
- [9] P. D. E. P. E. M. Aqüicultura et al., "PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN STRATEGI PROMOSI UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA YOGYAKARTA
- [10] Agus Nur Khormarudin, "Teknik Data Mining : Algoritma K-Means Clustering," pp. 1–12, 2016.
- [11] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018.
- [12] S. L. B. Ginting and R. P. Trinanda, "Teknik Data Mining Menggunakan Metode Bayes Classifier Untuk Optimalisasi Pencarian Pada Aplikasi Perpustakaan," *Univ. Pas.*, vol. d, no. Pencarian Informasi, pp. 1–14, 2015.
- [13] D. Yulianti, "Aplikasi Simulasi Laporan Keuangan Pribadi Berbasis Android," pp. 4– 16, 2017.
- [14] C. E. Grueber et al., "ANDROID DEVELOPER FUNDAMENTALS COURSE," in *Journal of Animal Ecology*, vol. 51, no. July, 2019, pp. 1– 12.
- [15] H. Saputro, "Pembelajaran Basis Data (Mysql)," *Modul Pembelajaran Prakt. Basis Data*, pp. 1–34, 2016.
- [16] R. V Palit, Y. D. Y. Rindengan, and A. S. M. Lumenta, "Rancangan Sistem Informasi Keuangan Gereja Berbasis Web Di Jemaat GMIM Bukit Moria Malalayang," *E-Journal Tek. Elektro Dan Komput.*, vol. 4, no. 7, pp. 1–7, 2015.
- [17] Y. Purbadian, "Framework Codeigniter 3," p. 144, 2016.
- [18] Codeigniter, "codeigniter." [Online]. Available: [https://codeigniter.com/user\\_guide/overview/appflow.html](https://codeigniter.com/user_guide/overview/appflow.html).
- [19] T. Rosandy, "PERBANDINGAN METODE NAIVE BAYES CLASSIFIER DENGAN METODE DECISION TREE (C4.5) UNTUK MENGANALISA KELANCARAN PEMBIAYAAN (Study Kasus : KSPPS / BMT AL-FADHILA)," *J. Teknol. Inf. Magister Darmajaya*, vol. 2, no. 01, pp. 52–62, 2016.
- [20] T. Chai, R. R. Draxler, and C. Prediction, "Root mean square error ( RMSE ) or mean absolute error ( MAE )? – Arguments against avoiding RMSE in the literature," no. 2005, pp. 1247– 1250, 2014