

SISTEM REKOMENDASI ARTIKEL BERITA MENGGUNAKAN METODE K-NEAREST NEIGHBOR BERBASIS WEBSITE

Sirajuddin Abraham¹⁾, Yeni Dwi Rahayu²⁾

^{1,2} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
email : mailsirogane@gmail.com, yenidwirahayu@unmuhjember.ac.id

Abstract

Based on data from the Ministry of Communication and Informatics (KOMINFO) in 2014 the number of internet users in Indonesia reached 88.3 million people. The figure puts Indonesia in the 6th position of internet users worldwide. It shows people have been using the internet in everyday life. The number of visitors on news websites can be utilized by managers and developers of a news site to attract visitors and access longer. Because based on statistics, website is often visited in Indonesia based alexa site is a search engine like google or yahoo and also social media like facebook. While the news site is Detik.com which ranks 9th. One way is to build a system of article recommender recommendations on a relevant website for users or visitors to obtain information. The goal of this recommendation system is also the owner of the website, so that website traffic and visitor statistics can go up so it can profit from managing the website. the accuracy of the error value in making the news recommendation system using K-nearest Neighbor method is 14%.

Keyword : Recommendation System, K-Nearest Neighbor, Article, News, Website.

1. PENDAHULUAN

1.1 Latar Belakang

Berdasarkan data Kementerian komunikasi Dan Informatika (KOMINFO) pada tahun 2014 jumlah pengguna internet di Indonesia mencapai 88,3 juta orang. Angka tersebut menempatkan negara Indonesia pada posisi ke-6 pengguna internet di seluruh dunia. Hal tersebut menunjukkan masyarakat telah menggunakan internet di kehidupan sehari-hari. Internet menawarkan kemudahan akses dalam mendapatkan informasi, salah satunya pada website berita di internet. Konsumsi akses berita telah mengalami perubahan dari model tradisional seperti koran, majalah dan media cetak lainnya menjadi akses ribuan berita dari berbagai sumber di internet.

Banyaknya pengakses pada website berita dapat dimanfaatkan oleh pengelola dan

pengembang sebuah situs berita untuk menarik pengunjung dan melakukan akses lebih lama. Karena berdasarkan statistik, website yang sering dikunjungi di Indonesia berdasarkan situs alexa adalah mesin pencari seperti google atau yahoo dan juga sosial media seperti facebook. Sedangkan situs berita adalah Detik.com yang menempati urutan 9.

Salah satu cara untuk membuat akses pengunjung lebih lama dari sebuah website adalah membuat sistem rekomendasi. Beberapa tahun terakhir sistem rekomendasi telah menjadi lazim digunakan untuk mengelola masalah data dan informasi (Wang et al. 2014). Sistem rekomendasi dapat memberikan saran atau pilihan artikel yang relevan untuk pengguna dalam mencari artikel berita pada sebuah website berita. Cabang ilmu pengetahuan yang dapat digunakan dalam membuat sistem rekomendasi salah satunya adalah data mining. Karena data

mining memiliki potensi besar dalam penggalan ilmu yang bermanfaat dari sejumlah besar data sementara untuk pengambilan keputusan yang dinamis (Ltifi et al. 2016).

Berdasarkan data dan uraian yang telah diuraikan dan penerapan metode KNN, maka penulis mencoba membuat sistem rekomendasi penentu artikel pada sebuah website yang relevan untuk pengguna atau pengunjung untuk memperoleh informasi. Sasaran dari sistem rekomendasi ini juga pemilik website tersebut, agar trafik website dan statistik pengunjung dapat naik sehingga dapat keuntungan dari mengelola website tersebut. Sistem yang akan dibuat dalam bentuk tugas akhir dengan judul “SISTEM REKOMENDASI PENENTU ARTIKEL PADA WEBSITE DENGAN MENGGUNAKAN METODE K-NEAREST NEIGHBOR.

1.2 Rumusan Masalah

Berdasarkan Latar belakang yang telah disusun, permasalahan yang akan diselesaikan adalah bagaimana membangun sebuah sistem rekomendasi untuk artikel sebuah website yang relevan dengan artikel yang telah diakses pengguna dalam waktu tertentu menggunakan metode K-Nearest Neighbor, tanpa harus menggunakan data profil pengguna, agar pengguna website memperoleh informasi yang lebih akurat.

1.3 Tujuan

Tujuan dari pembuatan tugas akhir ini adalah sebagai berikut :

- a. Membangun website yang akan digunakan sebagai media penyedia artikel dalam berbagai kategori
- b. Menerapkan metode *K-Nearest Neighbor* pada website yang telah dibangun untuk

mendapatkan keakuratan dari rekomendasi dari artikel yang telah diakses.

1.3 Manfaat

Manfaat yang diinginkan dicapai dari pembuatan tugas akhir ini adalah :

- a. Dapat mengimplementasikan teknik dari data mining yaitu *K- Nearest Neighbor* untuk penyelesaian masalah, dalam topik tugas akhir ini adalah membuat sistem rekomendasi.
- b. Sebagai penyedia website dapat memperoleh trafik pengunjung yang tinggi karena pengunjung dapat mengakses website lebih lama dan memperoleh pendapatan dari website tersebut lebih banyak.
- c. Sebagai pengguna website mendapat kemudahan memilih akses berita lain dari rekomendasi tersebut tanpa harus mencari berita-berita lain.

1.4 Batasan Masalah

Berdasarkan latar belakang yang telah disusun agar penyelesaian tugas akhir ini tidak menyimpang maka penulis menentukan batasan masalah sebagai berikut

- a. Pada tugas akhir ini penulis membuat *plugin* yang menjadi rekomendasi penentu artikel terkait berdasarkan artikel yang sedang diakses.
- b. Sistem rekomendasi untuk artikel adalah berdasarkan text mining dari judul, kategori artikel dari pengguna, yang dihasilkan dari metode *K-Nearest Neighbor*.
- c. Rekomendasi yang dihasilkan pada pembuatan tugas akhir ini tidak berdasarkan pada user profile, tetapi berdasarkan tagline berita yang sedang diakses

2. TINJAUAN PUSTAKA

2.1 Data Mining

Data mining didefinisikan sebagai kemampuan pencarian data yang canggih yang menggunakan algoritma statistik untuk menemukan pola dan korelasi dalam data (Rygielski et al. 2002). Teknik data mining telah menjadi cabang dari aplikasi kecerdasan buatan sejak tahun 1960. Selama dekade intervensi, inovasi penting dalam sistem komputer telah menyebabkan pengenalan dari teknologi yang baru untuk web-based education. Data mining mengizinkan mengizinkan pencarian untuk informasi yang bernilai dalam data yang memiliki volume besar (Liao et al. 2012).

2.2 K Nearest Neighbor

Algoritma K-Nearest Neighbor (KNN) adalah salah satu metode paling sederhana untuk memecahkan masalah klasifikasi; algoritma ini sering menghasilkan hasil yang kompetitif dan memiliki keuntungan yang signifikan selama beberapa metode data mining lainnya (Adeniyi Q2 et al. 2014). K-Nearest Neighbor (KNN) termasuk kelompok instance-based learning. Algoritma ini juga merupakan salah satu teknik lazy learning. KNN dilakukan dengan mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing (Wu, 2009). Ada banyak cara untuk mengukur jarak kedekatan antara data baru dengan data lama (data training), diantaranya euclidean distance dan manhattan distance (city block distance), yang paling sering digunakan adalah euclidean distance (Bramer, 2007).

2.3 Sistem Rekomendasi

Sistem rekomendasi didefinisikan sebagai strategi pengambilan keputusan untuk pengguna dibawah lingkungan informasi yang kompleks. Selain itu sistem rekomendasi

didefinisikan dari perspektif E-commerce adalah sebagai alat untuk membantu pengguna mencari melalui catatan pengetahuan yang berkaitan dengan minat dan preferensi pengguna. Sistem rekomendasi menangani masalah informasi yang berlebihan yang biasanya pengguna hadapi dengan cara personalisasi, konten eksklusif dan rekomendasi layanan (Isinkaye et al. 2015).

Sistem rekomendasi merupakan sebuah perangkat lunak yang bertujuan untuk membantu pengguna dengan cara memberikan rekomendasi kepada pengguna ketika pengguna dihadapkan dengan jumlah informasi yang besar. Rekomendasi yang diberikan diharapkan dapat membantu pengguna dalam proses pengambilan keputusan, seperti barang apa yang akan dibeli, buku apa yang akan dibaca, atau musik apa yang akan didengar, dan lainnya (Ricci et al, 2011).

2.4 Bahasa Pemrograman PHP

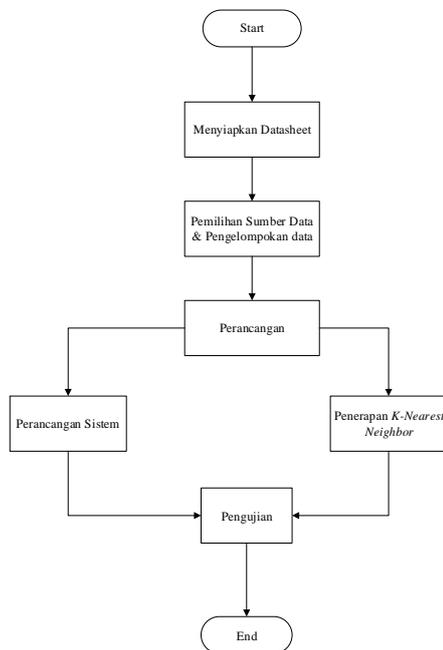
PHP adalah bahasa pemrograman script server-side yang didesain untuk pengembangan web. Selain itu, PHP juga bisa digunakan sebagai bahasa pemrograman umum. PHP dikembangkan pada tahun 1995 oleh Rasmus Lerdorf, dan sekarang dikelola oleh The PHP Group. PHP disebut bahasa pemrograman server side karena PHP diproses pada komputer server. Hal ini berbeda dibandingkan dengan bahasa pemrograman client-side seperti JavaScript yang diproses pada web browser (client).

3. METODE PENELITIAN

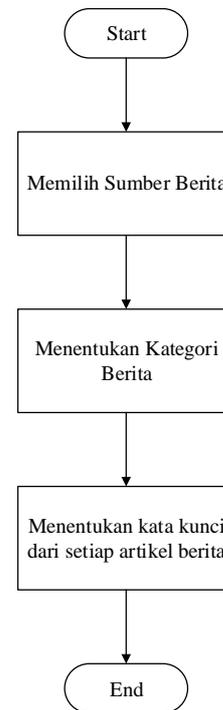
1. Menyiapkan Dataset

Pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan penelitian. Sebelum melakukan penelitian, seorang peneliti biasanya telah memiliki dugaan berdasarkan

teori yang ia gunakan, dugaan tersebut disebut dengan hipotesis.



Gambar 3.1 Alur Penelitian



Gambar 3.2 Pengelompokan Data

Menurut Sugiyono (2013:224) teknik pengumpulan data merupakan langkah yang paling strategis dalam penelitian, karena tujuan utama dari penelitian adalah mendapatkan data.

2. Pengumpulan dan Pengelompokan data

Proses dalam membuat sistem rekomendasi pada tugas akhir ini adalah menyiapkan data sebagai basis pengetahuan. Pada gambar 3.2 langkah awal dari pemilihan data untuk artikel berita adalah menentukan sumber yang akan digunakan untuk disimpan pada database dan ditampilkan pada halaman website. Untuk tugas akhir ini sumber berita yang akan digunakan adalah website dari metronews.com, Tempo.co dan vivanews.com. Rentang waktu artikel yang diambil dari tanggal 8 Januari 2017 sampai 20 januari 2017.

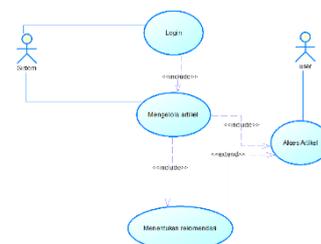
Tabel 3.1 jumlah berita yang diambil dari website

No	Nama Website	Jumlah berita per kategori		
		Ekonomi Bisnis	Dunia/Manca Negara	Sosial Politik
1	Metronews.com	20 Artikel	20 Artikel	20 Artikel
2	Tempo.co	20 Artikel	20 Artikel	20 Artikel
3	Vivanews.com	20 Artikel	20 Artikel	20 Artikel

3. Perancangan

Perancangan Sistem

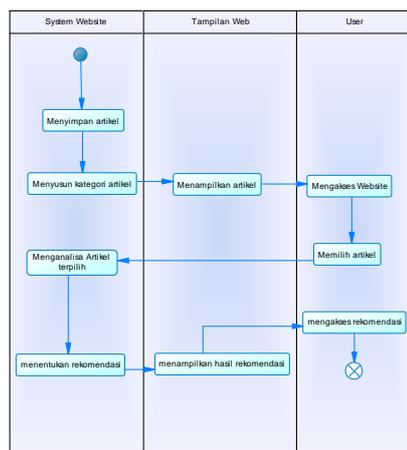
a. Diagram Usecase



Gambar 3.3 usecase diagram

Use case dari sistem yang akan dibuat terdiri dari aktor sistem dan user yang mengakses website yang berisi artikel dan sistem rekomendasi bagi user yang akan mengaksesnya. User mempunyai peran sebagai pengakses artikel. Semua artikel pada website dikelola oleh sistem yang disimpan pada storage, sistem mengelola artikel tersebut seperti mengedit, update atau bahkan menghapus isi dari artikel tersebut. Sedangkan untuk rekomendasi artikel yang akan disarankan kepada user, sistem menyipkan algoritma yaitu KNN agar sistem dapat menentukan artikel yang terkait secara otomatis.

b. Diagram Activity

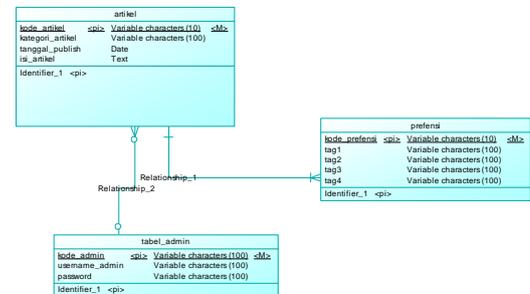


Gambar 3.4 Diagram Activity

Website yang akan dibangun memiliki sistem yang digunakan untuk menyimpan artikel dalam berbagai macam kategori. Selanjutnya artikel yang telah tersusun ditampilkan pada halaman website untuk diakses oleh pengguna. Setelah pengguna memilih artikel yang diinginkan, selanjutnya sistem menganalisa artikel yang dipilih untuk selanjutnya dilakukan perhitungan untuk mendapat artikel rekomendasi bagi user. Hasil

rekomendasi artikel berupa artikel serupa yang sedang diakses oleh user.

a. Desain Database



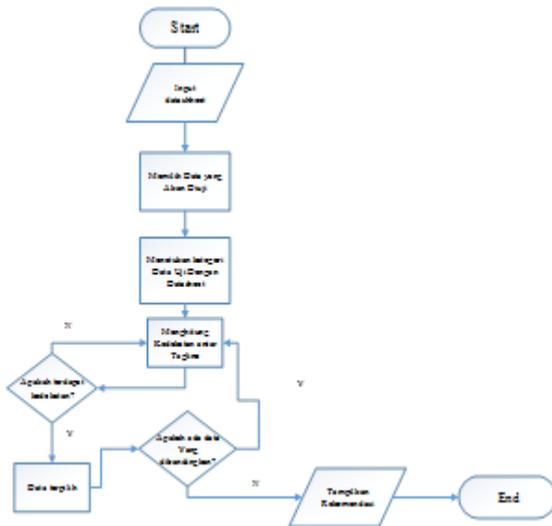
Gambar 3.5 Desain Database

Tabel database dalam perancangan sistem rekomendasi yang akan dibuat pada tugas akhir ini berjumlah 3 tabel, yang terdiri dari tabel artikel yaitu tabel yang berisi rincian dari artikel yang akan ditampilkan pada tampilan website. Kemudian tabel prefensi yang berisi tag atau tanda dari sebuah artikel yaitu rincian dari kata kunci atikel tersebut, yang disesuaikan dengan judul atau isi artikel. Selanjutnya tabel prefensi digunakan untuk data perhitungan dengan metode K-Nearest Neighbor. Tabel lain yang dibutuhkan adalah tabel admin yang bertugas mengupdate isi atrikel tersebut. relasi antar tabel untuk perancangan sistem rekomendasi tugas akhir ini terdapat seperti yang ditampilkan pada gambar.

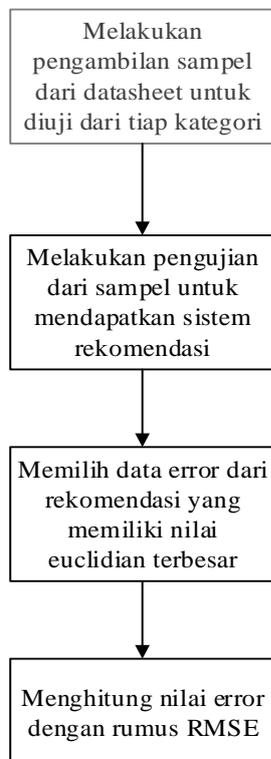
Penerapan metode KNN diawali dengan cara mengumpulkan datasheet yang diinput ke dalam database. Setelah proses input maka data ditampilkan dan user memilih data yang berbentuk artikel yang berada pada tampilan website. Data yang telah dipilih kemudian dibandingkan kedeketannya dengan kategori yang sama dengan data terpilih untuk memudahkan proses perhitungan. Data akan terus dibandingkan sampai memperoleh data

terdekat dengan metode perhitungan yang akan dipaparkan Pengujian

Penerapan KNN



Gambar 3.6 Penerapan KNN

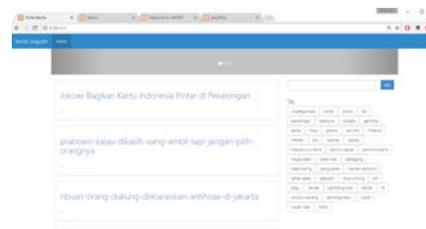


Gambar 3.7 Perhitungan Error

Pengujian dilakukan untuk melakukan analisa keakuratan menggunakan rumus akurasi dan menentukan error menggunakan rumus Root Mean Square Error (RMSE) dari penerapan metode K-Nearest Neighbor (KNN) untuk mendapatkan rekomendasi artikel berita yang akan dibangun. selain itu pengujian ini dilakukan untuk mengetahui apakah fitur-fitur yang ada pada sistem tersebut telah sesuai dengan kebutuhan pengguna sistem.

4. HASIL DAN PEMBAHASAN

4.1 Implementasi Desain Sistem



Gambar 4.1 Tampilan Website

Halaman depan dari website berita yang dibangun terdiri dari beberapa tampilan artikel berita yang nantinya diakses pengguna. Setelah pengguna memilih salah satu artikel maka sistem secara otomatis akan memberikan rekomendasi dari artikel yang telah dipilih. Berikut adalah tampilan ketika artikel dipilih dan rekomendasi yang dihasilkan.



Gambar 4.2 Penerapan KNN pada website

4.2 Penerapan Metode KNN

Penerapan KNN dilakukan untuk mendapatkan hasil rekomendasi berita dari

hasil perhitungan pada berita terpilih dengan membandingkan dengan dataset yang sudah ada. Pengujian dilakukan dengan beberapa skenario. Adapun skenario dalam pengujian metode KNN adalah sebagai berikut.

- A. Pengujian dilakukan melalui 5 kali iterasi di setiap kategori, sehingga total iterasi sebanyak 15 kali.
- B. Pemilihan artikel dilakukan secara acak pada masing-masing kategori
- C. Nilai kedekatan KNN dianggap ideal jika bernilai 2 hingga 4
- D. Nilai kedekatan dianggap data error jika bernilai 0 dan 1.
- E. Setelah artikel terpilih kemudian ditampilkan dengan nilai hasil perhitungan KNN

4.3 Analisa Pembahasan

Nilai KNN ideal untuk rekomendasi berita pada pembuatan tugas akhir ini adalah 2 hingga 4. Langkah selanjutnya adalah melakukan pencarian rekomendasi dengan nilai KNN 0 dan 1 yang dianggap sebagai data error. Langkah selanjutnya, menghitung jumlah total data yang error untuk mengetahui keakuratan metode KNN dalam menentukan rekomendasi dalam bentuk root mean square error (RMSE). Uji keakuratan dilakukan untuk mengetahui keefektifan metode KNN dalam menentukan rekomendasi artikel berita pada pembuatan tugas akhir ini. Setelah mengetahui presentase keakuratan, langkah selanjutnya adalah melakukan pengambilan keputusan, apakah metode KNN dapat dijadikan acuan dalam menentukan rekomendasi. Pengujian untuk mengukur keakuratan berdasarkan jumlah error pada masing-masing kategori. Berikut jumlah error berdasarkan kategori.

NO	Nama kategori	Jumlah Error
1.	Berita Politik	2
2.	Berita Internasional	11
3.	Ekonomi Bisnis	6

Pada pengujian kali ini, perhitungan dilakukan dengan cara mengolah nilai error masing-masing iterasi untuk mencari nilai Root Mean Square Error (RMSE). Adapun perhitungan RMSE adalah sebagai berikut.

$$RMSE = \sqrt{((2+11+6)/150)}^2 * 100\%$$

$$RMSE = 14\%$$

Maka nilai Error berdasarkan RMSE untuk menguji Metode perhitungan KNN dalam menentukan sistem rekomendasi artikel berita pada pembuatan tugas akhir ini adalah 14%.

5. KESIMPULAN DAN SARAN

Berdasarkan perancangan sistem dan analisa pada pembuatan tugas akhir ini, dapat ditarik kesimpulan sebagai berikut.

1. Metode KKN dapat digunakan sebagai sistem rekomendasi artikel berita berdasarkan tagline yang telah ditentukan dengan uji keakuratan berdasarkan nilai error sebesar 14%.
2. Setelah melakukan pengujian dengan beberapa iterasi, ditemukan artikel dengan rekomendasi yang tidak memenuhi syarat batas ketentuan metode KKN yang telah ditentukan.

Sistem Penentu Rekomendasi berita dengan menggunakan Metode KNN dapat dilakukan pengembangan lebih lanjut sehingga membutuhkan saran sebagai berikut.

1. Sistem rekomendasi ini menggunakan metode KNN dalam menentukan rekomendasi yang sesuai untuk pengguna. Maka sebaiknya dapat digunakan metode lain baik metode

pada data mining atau pun metode pada sistem rekomendasi ini. Baik menggunakan Content Based, Collaborative Filtering maupun metode Hybrid.

2. Pada perancangan tugas akhir ini, sistem rekomendasi yang dibangun tidak menggunakan data yang ada profil pengguna sebagai acuan untuk menghasilkan rekomendasi. Dalam pengembangan lebih lanjut, data user profil dapat digunakan sebagai penentuan rekomendasi sehingga bisa rekomendasi bisa diolah lagi.
3. Platform dalam pembuatan sistem rekomendasi bisa berbasis mobile dan data dapat menggunakan audio visual mengingat kebutuhan masyarakat beralih dari teks menjadi audio visual.

6. REFERENSI

- [1] Adeniyi Q2, D A, Z Wai, and Y Yongquan. 2014. Automated Web Usage Data Mining and Recommendation System Using KNearest Neighbor (KNN) Classification Method.
- [2] Berry, Michael J A, Gordon S Linoff, Markus Hegland, Mohammed J Zaki, and Limsoon Wong. 2003. Data Mining Techniques.
- [3] Brady M. Loonam J.2010. Exploring the use of entity-relationship diagramming as a technique to support grounded theory inquiry. Qualitative Research in Organization And Management
- [4] Herlawati& Widodo. 2011. Menggunakan UML. Informatika. Bandung
- [5] Injadat, Mohammadnoor, Fadi Salo, and Ali Bou Nassif. 2016. Data Mining Techniques in Social Media: A Survey
- [6] Isinkaye, F. O., Y. O. Folajimi, and B. A. Ojokoh. 2015. Recommendation Systems: Principles, Methods and Evaluation. Egyptian Informatics Journal 16 (3).
- [7] Jiang, Chuntao, Frans Coenen, and Michele Zito. 2004. A Survey of Frequet Subgraph Mining Algorithms. The Knowledge Engineering Review.
- [8] Kominformo.2014. Data pengguna internet (https://kominformo.go.id/content/detail/4286/pengguna-internet-indonesia-nomor-enam-dunia/0/sorotan_media) diakses : 9 Desember 2016.
- [9] Leidiyana, Henny. 2013. Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor. Jurnal Penelitian Ilmu Komputer, System Embedded & Logic.
- [10] Liao, Shu Hsien, Pei Hui Chu, and Pei Yuan Hsiao. 2012. Data Mining Techniques and Applications - A Decade Review from 2000 to 2011.
- [11] Liu, Jiahui, Peter Dolan, and Elin Rønby Pedersen. 2009. Personalized News Recommendation Based on Click Behavior.
- [13] Ltifi, Hela, Emna Benmohamed, Christophe [12] Kolski, and Mounir Ben Ayed. 2016. Enhanced Visual Data Mining Process for Dynamic Decision-Making. Knowledge-Based Systems. Elsevier B.V.

- [14] Nugroho.Adi. 2009. Rekayasa Perangkat Lunak Menggunakan UML & Java. Yogyakarta: Andi Offset.

- [15] Rygielski, Chris, Jyun-Cheng Wang, and David C. Yen. 2002. Data Mining Techniques for Customer Relationship Management.

- [16] Wang, Zan, Xue Yu, Nan Feng, and Zhenhua Wang. 2014. An Improved Collaborative Movie Recommendation System Using Computational Intelligence. *Journal of Visual Language and Computing* 25: 667675.

- [17] Yazdanfar, Nazpar, and Alex Thomo. 2013. LINK RECOMMENDER: CollaborativeFiltering for Recommending URLs to Twitter Users.