PENGKLASIFIKASIAN DATA SEKOLAH PENGGUNA INTERNET PENDIDIKAN MENGGUNAKAN TEKNIK *CLUSTERING* DENGAN ALGORITMA *K-MEANS* STUDI KASUS PT TELKOM SURABAYA

SKRIPSI

Diajukan Untuk Memenuhi Sebagian Syarat Guna
Memperoleh Gelar Sarjana Komputer (S.kom)
Pada Program Studi Teknik Informatika

OLEH :

VIVIN DWI RETNANINGSIH
NPM: 12.1.03.02.0417

FAKULTAS TEKNIK
UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA
UN PGRI KEDIRI
2016

Skripsi oleh:

VIVIN DWI RETNANINGSIH
NPM: 12.1.03.02.0417

Judul:

PENGKLASIFIKASIAN DATA SEKOLAH PENGGUNA INTERNET PENDIDIKAN MENGGUNAKAN TEKNIK CLUSTERING DENGAN ALGORITMA K-MEANS STUDI KASUS PT. TELKOM SURABAYA

Telah disetujui untuk diajukan kepada

Panitia Ujian/Sidang Skripsi Jurusan Teknik Informatika

Fakultas Teknik UN PGRI Kediri

Tanggal: 27 Juli 2016

Pembimbing I	Pembimbing II
Suratman, S.H., M.Pd.	**Ardi Sanjaya, M.Kom.**
NIDN. 0719036102	NIDN. 0706118101

Skripsi oleh:

VIVIN DWI RETNANINGSIH
NPM: 12.1.03.02.0417

Judul:

PENGKLASIFIKASIAN DATA SEKOLAH PENGGUNA INTERNET PENDIDIKAN MENGGUNAKAN TEKNIK CLUSTERING DENGAN ALGORITMA K-MEANS STUDI KASUS PT. TELKOM SURABAYA

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi
Program Studi Teknik Informatika Fakultas Teknik UNPGRI Kediri
Pada tanggal: 12 Agustus 2016

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Penguji:

1. Ketua : Suratman, SH., M.Pd _____
2. Penguji I : Rini Indriati, S.Kom., M.Kom. _____
3. Penguji II : Ardi Sanjaya, M.Kom. _____

Mengetahui,
Dekan Fakultas Teknik

Dr. Suryo Widodo, M.Pd
NIP.19640202 199103 1 002

PENGKLASIFIKASIAN DATA SEKOLAH PENGGUNA INTERNET PENDIDIKAN MENGGUNAKAN TEKNIK CLUSTERING DENGAN ALGORITMA K-MEANS STUDI KASUS PT. TELKOM SURABAYA

Vivin Dwi Retnaningsih
12.1.03.02.0417
Teknik – Teknik Informatika
vivinreparen@gmail.com
Suratman, SH., M.Pd dan Ardi Sanjaya, M.Kom
UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Vivin Dwi Retnaningsih : Pengklasifikasian Data Sekolah Pengguna Internet Pendidikan Menggunakan Teknik *Clustering* dengan Algoritma *K-Means* Studi Kasus PT. Telkom Surabaya, Skripsi, Teknik Informatika, Fakultas Teknik UN PGRI Kediri, 2016.

Penelitian ini dilatar belakangi hasil pengamatan peneliti, bahwa banyaknya data yang dipunyai sebuah instansi bisa menyebabkan kesulitan dalam mengelola data tersebut untuk kepentingan instansi. Akibatnya banyak terjadi kesalahan dalam pengkasifikasian yang dilakukan.

Permasalahan penelitian ini adalah bagaimana cara merancang sistem untuk pengolahan data sekolah dengan menggunakan metode *clustering*?

Di dalam penelitian ini menggunakan pendekatan Penelitian Rekayasa Perangkat Lunak dengan subyek penelitian sekolah pengguna internet pendidikan di PT. Telkom Surabaya. Penelitian ini menggunakan algortima *k-means* dalam mengelompokkan data sekolah. Data sekolah dikelompokkan berdasarkan jumlah kluster yang ditentukan. Selanjutnya, setiap data/obyek ditempatkan berdasarkan kluster terdekat menggunakan konsep jarak *euclidean distance*. Proses tersebut terus dilakukan sampai pusat kluster tidak berubah lagi.

Hasil penelitian ini adalah (1) Melalui algoritma *K-Means* ini terbukti dapat membantu mengelompokan data sekolah dalam jumlah besar berdasarkan kluster yang telah ditentukan secara *random* (acak). (2) Dengan menggunakan metode *Clustering* dengan algoritma *K-Means* terbukti dapat mengelola dan mengelompokkan data sekolah serta tagihan sekolah secara otomatis.

Berdasarkan hasil penelitian ini, direkomendasikan : (1) Tujuan pokok penggunaan algoritma *k-means* ini adalah untuk mengelompokkan data sekolah. Oleh sebab itu karyawan PT. Telkom Surabaya sebagai pengguna aplikasi ini harus mengutamakan proses yang mendukung kelancaran penggunaan dan perawatan aplikasi ini. (2) Karyawan PT. Telkom Surabaya perlu mengembangkan lagi aplikasi yang telah dibuat oleh peneliti. (3) Sekolah-sekolah di Surabaya terutama pengguna aplikasi inidiharapkan bisa lebih mengembangkan aplikasi ini dengan lebih baik dan lebih bermanfaat lagi.

Kata kunci : Pengklasifikasian, Data Mining, *Clustering*, *K-Means*, *Euclidean Distance.*

I. LATAR BELAKANG

Banyaknya data yang dipunyai sebuah instansi bisa menyebabkan kesulitan dalam mengelola data tersebut untuk kepentingan instansi. Kegiatan pengklasifikasian yang dilakukan oleh manusia masih memiliki keterbatasan, terutama pada kemampuan manusia dalam menampung jumlah data yang ingin dikelola. Selain itu bisa juga terjadi kesalahan dalam pengklasifikasian yang dilakukan. Salah satu cara mengatasi masalah ini adalah dengan menggunakan *Data Mining* (DM) dengan teknik *clustering*.

Penggunaan teknik DM *clustering* berbeda–teknik *Data Mining* (DM) yang lainnya, seperti *association rule mining* dan *classification* yang memerlukan tahapan *training* dan evauasi. Teknik ini menggunakan metode *unsupervised learning* yang berarti DM tidak perlu melakukan *training* terlebih dahulu tapi bisa langsung menggunakannya untuk pengelompokkan. Teknik ini masih jarang digunakan dibanding dengan teknik-teknik DM yang lain.

Data Mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Patut diingat bahwa kata *mining* sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu DM sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistic dan basis data. Beberapa teknik yang sering disebut-sebut dalam *literatur* DM antara lain :*clustering, classification, association rule mining, neural network,* dan *genetic algorithm*.

PT. Telkom Surabaya adalah salah satu instansi milik BUMN yang terbesar di Jawa Timur. Dalam perkembangannya PT. Telkom Surabaya banyak mengeluarkan produk-produk yang bermanfaat terutama bagi warga Surabaya dan warga seluruh Indonesia maupun manca negara. Salah satu produk yang saat ini digencarkan adalah produk *Indihome* dan Internet Pendidikan.

PT. Telkom Surabaya bekerja sama dengan pemerintah kota Surabaya untuk mewajibkan seluruh sekolah di Surabaya baik SD, SMP, maupun SMA untuk menggunakan Internet Pendidikan guna informasi tentang dunia pendidikan dapat tersebar kepada para siswa dengan lebih cepat dan akurat lagi, tanpa harus ada gangguan jaringan.

Banyaknya data sekolah yang ada di Surabaya membuat PT. Telkom Surabaya sulit untuk mengelola dan mengklasifikannya, selain itu banyak data sekolah yang tidak valid, sehingga membuat PT. Telkom Surabaya harus lebih ekstra lagi dalam memvalidkan data-data sekolah tersebut. Kegiatan klasifikasi yang dilakukan oleh manusia masih memiliki keterbatasan,

terutama pada kemampuan manusia dalam menampung jumlah data yang ingin diklasifikasikan. Selain itu bisa juga terjadi kesalahan dalam pengklasifikasian yang dilakukan.

Karena kurang validnya data sekolah yang diberikan oleh Pemkot Surabaya dapat mempengaruhi strategi promosi yang dilakukan oleh PT. Telkom Surabaya.

Menurut Lindawati, dalam jurnalnya yang berjudul "DATA MINING DENGAN TEKNIK *CLUSTERING* DALAM PENGKLASIFIKASIAN DATA MAHASISWA STUDI KASUS PREDIKSI LAMA STUDI MAHASISWA UNIVERSITAS BINA NUSANTARA", mampu menjadi referensi saya dalam menyusun strategi pemecahan masalah klasifikasi dan pengelolaan data yang dialami oleh PT. Telkom Surabaya. Teknik *clustering* efektif untuk mengklasifikasikan prediksi lama studi mahasiswa dengan menggunakan parameter-parameter tertentu.

Menurut Rima Dias Ramadhani, dalam jurnalnya yang berjudul "DATA MINING MENGGUNAKAN ALGORITMA *K-MEANS CLUSTERING* UNTUK MENENTUKAN STRATEGI PROMOSI UNIVERSITAS DIAN NUSWANTORO", mampu membantu saya dalam perancangan strategi promosi untuk produk Internet Pendidikan yang dialami oleh PT. Telkom Surabaya. Salah satu cara mengatasi masalah ini adalah dengan menggunakan *Data Mining* (DM) dengan teknik *clustering.*

Menurut Ferlyna K Wardhani, Erma Suryani, dan Ahmad Mukhlason, dalam jurnalnya yang berjudul "PENERAPAN METODE GA - *KMEANS* UNTUK PENGELOMPOKAN PENGGUNA PADA BAPERSIP PROVINSI JAWA TIMUR", mampu membantu saya dalam menyusun strategi pemecahan masalah pengelompokan data paket *bandwith* sekolah pengguna internet pendidikan pada PT Telkom Surabaya. Algoritma *K-means* mengelompokan data berdasarkan klaster-klaster yang telah ditentukan sebelumnya. Selain itu, algoritma GA-*Kmeans* mampu menghasilkan pengelompokan dengan tingkat kerapatan di dalam klaster yang lebih baik dibandingkan dengan algoritma *K-means* sederhana.

Pengolahan data sekolah pengguna internet pendidikan di kota Surabaya seharusnya dilakukan dengan tepat dan cepat, agar tidak terjadi kesalahan pemasangan kecepatan akses internet pendidikan.

Penelitian ini akan melakukan pengelompokan sesuai kecepatan akses internet yang dipesan oleh masing-masing sekolahan. Selain itu penelitian ini juga dapat membantu strategi promosi internet pendidikan oleh PT. Telkom Surabaya.

Dalam penelitian ini, saya menggunakan metode *clustering* dalam menyusun sistem pengolahan data. Data akan secara otomatis diolah oleh sistem, sehingga PT. Telkom dapat secara cepat mendapat informasi dan melakukan pemasangan secara tepat.

II. METODE

Metode *K-Means* bisa menggunakan ukuran kemiripan untuk mengelompokkan suatu obyek. Kemiripan ini diterjemahkan dalam konsep jarak (Santoso, 2007). Semakin dekat jarak, maka semakin tinggi kemiripannya.

Berikut langkah-langkah algoritma *K-Means Clustering* :

1. Menentukan jumlah K *cluster*.
2. Inisialisasi nilai K pusat *cluster* atau nilai *means* atau *centroid*, yang paling sering digunakan adalah dengan cara *random* (acak).
3. Menempatkan setiap data/obyek ke *cluster* terdekat. Kedekatan dua obyek ditentukan berdasar jarak kedua obyek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster*. Konsep jarak yang digunakan adalah jarak *Euclidean* (L_2-norm).
4. Menghitung kembali pusat *cluster* dengan keanggotaan *cluster* sekarang. Pusat *cluster* adalah rata-rata dari semua data atau obyek dalam *cluster* tertentu. Jika dikehendaki bisa juga memakai median dari *cluster* tersebut.
5. Menghitung kembali pusat *cluster* yang baru jika pusat tidak berubah lagi, maka proses pengklasteran selesai atau kembali ke langkah yang ketiga sampai pusat klaster tidak berubah lagi.

Hasil klaster dengan metode *K-Means Clustering* sangat bergantung pada nilai pusat klaster awal yang diberikan. Pemberian nilai awal yang berbeda bisa menghasilkan hasil klaster yang berbeda. Ada beberapa cara member nilai awal misalnya dengan mengambil sampel awal dari data atau member nilai awal secara *random* (acak).

Beberapa permasalahan yang sering muncul pada saat menggunakan metode *K-Means* untuk melakukan pengelompokkan data adalah :

a. Ditemukannya beberapa model klastering yang berbeda.
b. Pemilihan jumlah klaster yang paling tepat.
c. Kegagalan untuk *converge*.

d. Pendeteksian *outlier*.
e. Bentuk masing-masing klaster.
f. Masalah *overlapping*.

Keenam permasalahan diatas adalah beberapa hal yang perlu di perhatikan pada saat menggunakan *K-Means* dalam mengelompokkan data. Permasalahan 1 umumnya disebabkan oleh perbedaan proses inisialisasi anggota masing-masing klaster. Proses initialisasi yang sering digunakan adalah proses inisialisasi secara *random*.

Permasalahan 2 merupakan masalah laten dalam metode *K-Means*. Beberapa pendekatan telah digunakan dalam menentukan jumlah klaster yang paling tepat untuk suatu dataset yang dianalisa termasuk diantaranya *Partition Entropy* (PE) dan GAP *Statistics*. Metode-metode tersebut tidak sama dengan pendekatan yang digunakan oleh *K-Means* dalam mempartisi data items ke masing-masing klaster.

Permasalahan kegagalan untuk *converge*, secara teori memungkinkan untuk terjadi dalam kedua metode *K-Means* yang dijelaskan di dalam tulisan ini. Kemungkinan ini akan semakin besar terjadi untuk metode *Hard K-Means*, karena setiap data di dalam dataset dialokasikan secara tegas (*hard*) untuk menjadi bagian dari suatu klaster tertentu. Kejadian seperti ini tentu akan mengakibatkan pemodelan tidak akan berhenti dan kegagalan untuk *converge* akan terjadi. Untuk *K-Means*, walaupun ada, kemungkinan permasalahan ini untuk terjadi sangatlah kecil, karena setiap data dilengkapi dengan *membership function* untuk menjadi anggota klaster yang ditemukan.

Permasalahan keempat merupakan permasalahan umum yang terjadi hampir disetiap metode yang melakukan pemodelan terhadap data. Khusus untuk metode *K-Means* hal ini memang menjadi permasalahan yang cukup menentukan. Beberapa hal yang perlu diperhatikan dalam melakukan pendeteksian *outlier* dalam proses pengelompokkan data termasuk bagian menentukan apakah suatu data item merupakan *outlier* dari suatu klaster tertentu dan apakah data dalam jumlah kecil yang membentuk suatu klaster tersendiri dapat dianggap sebagai *outlier*. Proses ini memerlukan suatu pendekatan khusus yang berbeda dengan proses pendeteksian *outlier* di dalam suatu dataset yang hanya terdiri dari satu populasi yang homogen.

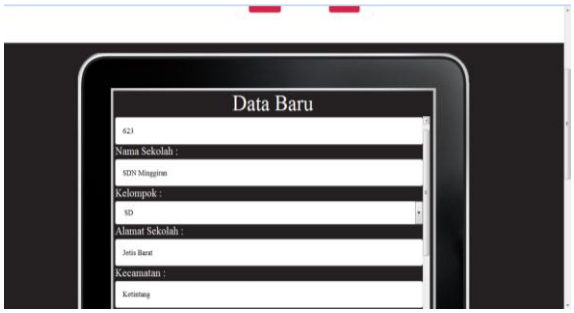
Permasalahan kelima adalah menyangkut bentuk klaster yang ditemukan. Tidak seperti metode data *clustering* lainnya termasuk *Mixture Modelling*, *K-Means* umumnya tidak mengindahkan bentuk dari masing-masing klaster yang mendasari model yang

terbentuk, walaupun secara natural masing-masing klaster umumnya berbentuk bundar. Untuk *dataset* yang diperkirakan mempunyai bentuk yang tidak biasa, beberapa pendekatan perlu untuk diterapkan.

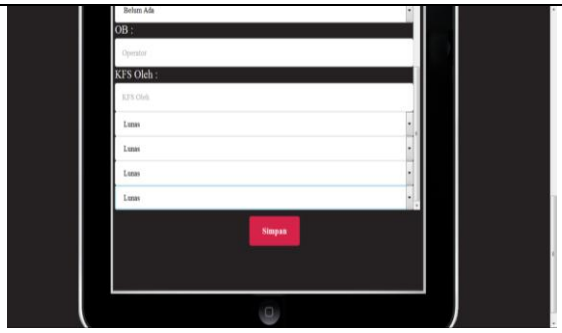
Masalah *over lapping* sebagai permasalahan terakhir sering sekali diabaikan karena umumnya masalah ini sulit terdeteksi. Hal ini terjadi untuk metode *Hard K-Means* dan *Fuzzy K-Means*, karena secara teori metode ini tidak diperlengkapi *feature* untuk mendeteksi apakah didalam suatu klaster ada klaster lain yang kemungkinan tersembunyi. (Yunus Ariadi Saputra, 2014)

III. HASIL DAN KESIMPULAN

Didalam menu data terdapat submenu tambah data, olah data, dan tagihan pelanggan. Berikut adalah tampilan dari form tambah data dan form olah data :



Gambar 5.10 Form Tambah Data



Gambar 5.11 Form Tambah Data Tagihan

Setelah data selesai di inputkan, klik simpan. Kemudian akan muncul gambar 5.12 seperti dibawah ini, masukkan nama sekolah yang baru diinputkan, klik hitung status untuk menghitung status berlangganannya.

Gambar 5.12 Form Hitung Status *K-Means*

Kemudian klik cari untuk mengetahui status berlangganan dari sekolah tersebut, seperti gambar dibawah ini :

Gambar 5. 13 Form Status Berlangganan

KESIMPULAN

1. Melalui algortima *K-Means* ini terbukti dapat membantu mengelompokkan data sekolah dalam jumlah besar berdasarkan kluster yang telah ditentukan secara *random* (acak).
2. Dengan menggunakan metode *Clustering* dengan algoritma *K-Means* terbukti dapat mengelola dan mengelompokkan data sekolah serta tagihan sekolah secara otomatis.

IV. DAFTAR PUSTAKA

Ariadi, Yunus S. 2014. Analisis Data Mahasiswa Menggunakan Algoritma *K- Means Clustering* (Studi Kasus Prodi Teknik Informatika Fakultas Teknik UNP Kediri). Universitas Nusantara PGRI Kediri

Budiman, Irwan. 2012. Data *Clustering* Menggunakan Metodologi *CRISP-DM* Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma. Universitas Diponegoro Semarang Diunduh pada 17 November 2015

Kadir, Abdul. 2013. *Pengenalan Sistem Informasi*. Yogyakarta : Andi

Permana, Galih. 2013. Pengertian PHP dan Kelebihannya (online). Tersedia : http://infoter-lengkap.blogspot.com/2013/03/pengertian-php-dan-kelebihannya.html Diunduh pada 17 November 2015

Portal koporet PT Telekomunikasi Indonesia, Tbk. 2015. http://www.telkom.co.id Diunduh pada 17 November 2015

Santosa, B. 2007. *Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Yogyakarta : Graha Ilmu

Suprianto, Dodit. 2008. *Buku Pintar Pemrograman PHP*. Jakarta : Oase Media

Yakub. 2002. *Pengantar Sistem Informasi* (Edisi 1). Yogyakarta : Graha Ilmu