

**DEVELOPMENT OF DATA WAREHOUSE TO PREDICATE THE REGARDING OF
UMROH CONGREGATIONS USING THE MANEAREST NEIGHBOUR ALGORITHM
(CASE STUDY PT. BAHANA SUKSES SEJAHTERA)**

Akhmad Wijaya Kusuma^{*1}, Deni Mahdiana²

^{1,2}Magister Ilmu Komputer, Fakultas Teknologi Informasi Universitas Budi Luhur, Indonesia
Email: ¹akhmadwijaya.kusuma@gmail.com, ²deni.mahdiana@budiluhur.ac.id

(Naskah masuk: 5 April 2022, Revisi: 11 April 2022, diterbitkan: 20 Agustus 2022)

Abstract

PT. Bahana Sukses Sejahtera is a travel company whose one goal is to serve Umrah activities to the holy land of Mecca, Saudi Arabia. The company has been carrying out travel activities for 15 years. Interest from travel activities, especially with the aim of Umrah to the holy land of Mecca from year to year is increasing because the economic conditions of prospective pilgrims are getting better. More and more Umrah travel enthusiasts certainly create abundant data but of course if not processed properly the data will not produce good information. How to get good information with good data processing so that information can be used as knowledge for company decision making. The activity of utilizing big data into information is to use information technology and databases and then design a data warehouse. Data warehouse design and data mining is the most effective technique to produce information that has knowledge for corporate decision making. The knowledge generated from the data warehouse is that one of them can produce information to predict prospective Umrah travel pilgrims who may be canceled due to something that makes it impossible to leave. One of the algorithms used for the predicting process can use the nearest neighbor algorithm. The results of the algorithm can produce predictive information about factors that are very influential on the cancellation of the departure of prospective Umrah pilgrims, for example due to family reasons. Based on this explanation, this research was developed by utilizing information technology and software to make predictive analysis based on existing data regarding the cancellation of the departure of Umrah pilgrims at the travel company PT. Bahana Sukses Sejahtera.

Keywords: data mining, data warehouse, informasi, travel umroh, nearest neighbour.

**PENGEMBANGAN DATA WAREHOUSE UNTUK MEMPREDIKSI
PENGUNDURAN DIRI JEMAAH UMROH MENGGUNAKAN
ALGORITMA NEAREST NEIGHBOUR
(STUDI KASUS PT. BAHANA SUKSES SEJAHTERA)**

Abstrak

PT. Bahana Sukses Sejahtera adalah salah satu perusahaan bidang travel yang salah satu tujuannya adalah melayani kegiatan umroh ke tanah suci Mekkah Saudi Arabia. Perusahaan sudah melaksanakan kegiatan travel sudah dilaksanakan dalam 15 tahun. Peminat dari kegiatan travel terutama dengan tujuan umroh ke taah suci Mekkah dari tahun ke tahun semakin meningkat dikarenakan kondisi ekonomi calon jamaah semakin baik. Peminat perjalanan umroh semakin lama semakin banyak tentunya menciptakan data yang berlimpah tetapi tentunya jika tidak diolah dengan baik data tersebut tidak menghasilkan informasi yang baik. Cara mendapatkan informasi yang baik dengan pengolahan data yang baik sehingga informasi tersebut dapat digunakan sebagai pengetahuan untuk pengambilan keputusan perusahaan. Kegiatan pemanfaatan data besar menjadi informasi adalah dengan menggunakan teknologi informasi dan database kemudian merancang sebuah data *warehouse*. Perancangan data *warehouse* dan data *mining* merupakan teknik paling efektif menghasilkan informasi yang memiliki pengetahuan untuk pengambilan keputusan perusahaan. Pengetahuan yang dihasilkan dari data *warehouse* tersebut adalah salah satunya dapat menghasilkan informasi untuk memprediksi calon Jemaah travel umroh yang mungkin akan batal berangkat dikarenakan sesuatu hal yang tidak memungkinkan untuk berangkat. Salah satu algoritma dipakai untuk proses memprediksi dapat menggunakan algoritma *nearest neighbour*. Hasil dari algoritma tersebut dapat menghasilkan informasi prediksi tentang faktor yang sangat berpengaruh terhadap batalnya keberangkatan calon Jemaah umroh contohnya adalah karena alasan keluarga. Berdasarkan penjelasan tersebut penelitian ini dikembangkan dengan memanfaatkan teknologi informasi dan perangkat lunak untuk

membuat analisa prediksi berdasarkan data-data yang ada tentang batalnya keberangkatan peserta Jemaah umroh di perusahaan travel PT. Bahana Sukses Sejahtera.

Kata kunci: *data mining, data warehouse, informasi, travel umroh, nearest neighbour.*

1. PENDAHULUAN

Seiring dengan perkembangan ekonomi dimasyarakat yang saat ini terus berkembang minat terhadap perjalanan spiritual semakin banyak diminati. Daerah perkotaan seperti di Propinsi Banten dan khususnya di Daerah Tangerang sekitarnya banyak masyarakat berminat terhadap perjalanan spiritual seperti umroh dan haji ke tanah suci Mekkah Saudi Arabia. Berdasarkan kondisi tersebut tercatat setiap tahun minat terhadap perjalanan umroh semakin meningkat tentunya secara otomatis meningkatkan jumlah *database* Jemaah umroh tersebut dalam sistem informasi perusahaan penyedia layanan perjalanan umroh dan haji.

Penumpukan data-data tersebut jika tidak dimanfaatkan secara baik terbuang secara sia-sia. Penumpukan data-data tersebut jika diolah dan dimanfaatkan dengan baik dan benar dapat meningkatkan kemajuan bagi perusahaan perjalanan umroh dan haji tersebut. PT. Bahana Sukses Sejahtera merupakan perusahaan yang bergerak dalam pelayanan perjalanan spiritual seperti umroh dan haji sebaiknya secara cermat memanfaatkan data-data tersebut menjadi lebih berguna sebagai pengembangan perusahaan. Perusahaan perjalanan yang berlokasi di Ruko Dasana Indah Blok SA No. 32 Bojong Nangka Kelapa Dua Tangerang Banten 15821 saat ini belum secara optimal memanfaatkan teknologi informasi tersebut untuk mengolah data menjadi informasi menambah pengetahuan pengembangan perusahaan.

Pengembangan sistem informasi yang terdapat pada database untuk menjadi informasi yang bernilai pengetahuan penunjang keputusan perusahaan adalah dengan membuat data *warehouse*. Perancangan data *warehouse* menjadi solusi pilihan dengan memanfaatkan data-data yang tersimpan didalam database besar yang dapat menghasilkan berbagai macam informasi yang lebih bermanfaat untuk perusahaan. Model Data *warehouse* menjadi salah satu metode memanfaatkan perangkat lunak database, untuk mendukung Sistem Penunjang Keputusan dan Sistem Informasi Eksekutif. Pengambilan keputusan perusahaan dapat dilaksanakan dengan mengambil data-data dari berbagai dimensi hasil pengolahan *database* besar tersebut.

Proses pengembangan manajemen database dengan model data *warehouse* kemudian dapat diteruskan ke pengembangan data berikutnya dengan menerapkan model data mining. Proses pengolahan data dengan cara menggali dari lebih lanjut untuk menghasilkan informasi yang mendukung

pengetahuan untuk membantu dalam pengambilan keputusan merupakan proses dari kegiatan Data Mining. Proses pengembangan Data *Warehouse* dan Data *Mining* ini memanfaatkan teknologi Kecerdasan buatan atau *Artificial Intelligence*, kemudian ilmu statistik dan ilmu matematika. Teknologi manajemen pengolahan data dengan fungsi penggalian data dalam dapat memberikan informasi tambahan dalam memberikan pendapat yang bersifat ilmiah dalam menghasilkan keputusan yang tepat suatu kegiatan bisnis.

Sesuai dengan penjelasan pada uraian paragraf sebelumnya penelitian yang dilakukan ini dikembangkan untuk membantu perusahaan PT. Bahana Sukses Sejahtera dalam mengembangkan perusahaan. Pengembangan perusahaan ini dimulai dengan memanfaatkan data-data yang tersimpan dalam database sistem informasi perusahaan menjadi sesuatu yang bermanfaat dalam pengambilan keputusan. Salah satu informasi dalam bentuk pengetahuan dari pengolahan *database* kemudian menjadi data *warehouse* dan data mining adalah fungsi prediksi atau perkiraan berdasarkan data. Salah satu prediksi yang dapat dikembangkan adalah tentang pembatalan calon Jemaah umroh karena alasan yang berbeda-beda. Tema inilah menjadi bahasan dalam penelitian ini dengan memanfaatkan teknologi informasi.

2. METODE PENELITIAN

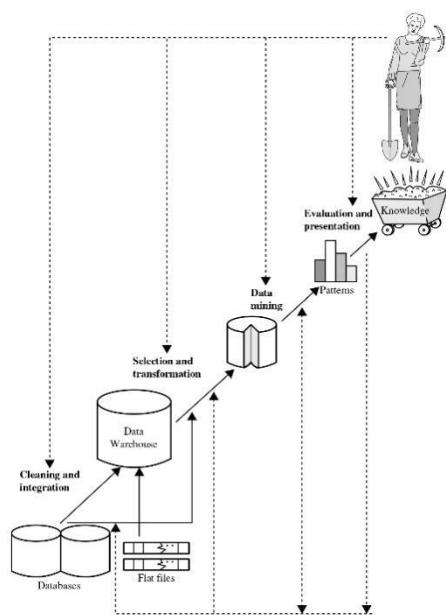
2.1 Data Warehouse

Data *warehouse* merupakan kumpulan data-data saling terhubung dan memiliki sifat berorientasi objek, *time variant*, tidak tetap untuk membantu pihak manajemen dalam mengambil keputusan [1]. Pengertian lain dari Data *warehouse* adalah sekumpulan data-data dari banyak sumber data kemudian diletakan dalam satu database selanjutnya dilakukan pengolahan ke bentuk *database* multi-dimensional untuk memudahkan pembuatan perintah dan laporan [2].

Data *warehouse* memiliki karakteristik berorientasi subjek, setiap data saling terhubung, memiliki jangka waktu dan bersifat tetap [3]. *Dimensionality modelling* merupakan teknik desain secara logika untuk mendapatkan data dalam bentuk standar dan intuitif bertujuan agar proses akses database memiliki performa yang bagus [1]. Model data skema multidimensioanl terdapat tipe-tipe seperti *stars, fact constellations, dan snowflakes* [4].

2.2 Data Mining

Model pengolahan data dengan menganalisa dari sekumpulan data-data sehingga mendapatkan hubungan data yang tidak diperkirakan sebelumnya kemudian meringkas data dengan teknik berbeda sehingga dapat dipahami kemudian dimanfaatkan lagi secara lebih baik oleh pengguna merupakan proses dari *Data Mining* [5]. Pencarian pola dengan teknik tertentu untuk mendapatkan informasi menarik merupakan pengertian lain dari *Data mining*. *Data mining* memiliki metode, teknik, dan algoritma sangat bervariasi. Keseluruhan metode yang digunakan atau algoritma yang digunakan dalam *data mining* dipengaruhi oleh proses dan fungsi *Knowledge Discovery in Database* [6].



Gambar 1. Proses kegiatan Data Mining

Terdapat 2 kelompok dalam *data mining* yaitu sebagai fungsi minor dan fungsi *major* [6]. Fungsi yang bekerja dengan cara mendeskripsikan, membuat estimasi dan memprediksi merupakan contoh bentuk fungsi minor. Fungsi mayor dalam *data mining* bisa dalam bentuk proses klasifikasi data, asosiasi data dan penggolongan data.

2.3 Algoritma *nearest Neighbor*

Algoritma *Nearest Neighbour* bekerja dengan cara mencari perhitungan pendekatan antar kasus baru dengan kasus sebelumnya berdasarkan jumlah fitur dan bobot [6]. Perhitungan kedekatan tersebut memiliki nilai nol sampai satu. Jika bernilai 0 berarti kedua kasus tidak memiliki kemiripan. Seandainya memiliki nilai 1 maka antar kasus tersebut mutlak memiliki kemiripan yang sama [6].

2.4 WEKA

Weka merupakan perangkat lunak yang dapat digunakan dalam proses pengolahan *data mining* dimana perangkat ini berbasis open source dibangun menggunakan bahasa pemrograman java [2]. Universitas Waikato dari Selandia Baru merupakan

institusi yang mengembangkan aplikasi ini dimana selanjutnya menjadi bagian dari perangkat lunak Pentaho. Perangkat lunak Weka memiliki beberapa algoritma sebagai *machine learning* untuk melakukan proses generalisasi rumus pengolahan data-data sampling.

3. HASIL DAN PEMBAHASAN

3.1 Profil PT. Bahana Sukses Sejahter

Perusahaan PT. Bahana Sukses Sejahtera memiliki alamat kantor di Ruko Dasana Indah Blok SA No. 32 Bojong Nangka Kelapa Dua Tangerang Banten 15821. Perusahaan penyedia layanan jasa perjalanan umroh ini telah berdiri selama 15 tahun dan sudah melaksanakan pengiriman Jemaah umroh ke tanah suci Mekkah Saudi Arabia. Paket perjalanan umroh ini juga terdapat paket tambahan perjalanan ke wisata spiritual lain di sekitar jajirah Arab.

3.2 Perancangan Data Warehouse

Model proses pengembangan *data warehouse* terdapat tahapan-tahapan yang dikenal dengan sembilan tahapan dimana tahapan-tahapan tersebut seperti berikut ini [1]:

a. Pemilihan Proses

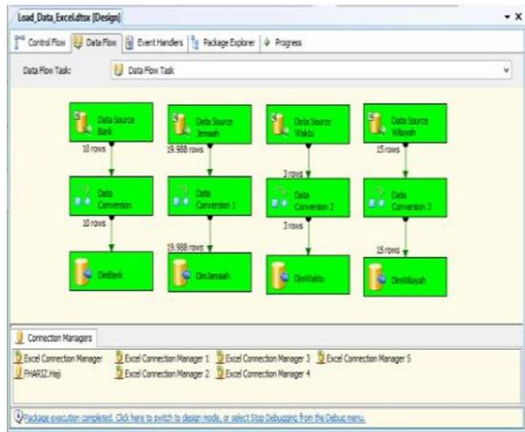
Data warehouse tentang proses keberangkatan Jemaah umroh terdiri dari Jemaah yang siap berangkat dan Jemaah yang batal berangkat. Atribut data yang ada kelompok jemaah, terdiri dari nama calon jemaah umroh, periode tahun keberangkatan, jenis kelamin jemaah dan rekening bank tabungan.

b. Pemilihan Grain

Model dimensi data berbentuk tabel fakta terdiri dari data bank, data jemaah, data waktu dan data wilayah. Perancangan *data warehouse* dalam grain PT. Bahana Sukses Sejahtera meliputi jumlah jemaah umroh dengan kelompok sesuai bank, jumlah jemaah sesuai daerah kabupaten kecamatan, sedangkan proses analisis digunakan tahun sebagai periode waktu.

c. Identifikasi dan Penyesuaian Dimensi

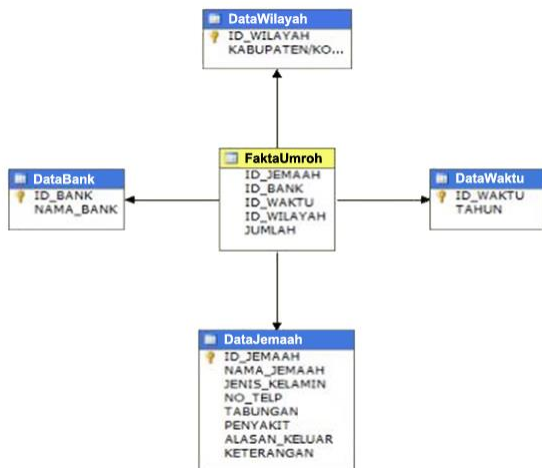
Tahapan ini dilakukan proses penyesuaian dimensi berdasarkan data-data tabel fakta.



Gambar 2. Loading Data pada SSIS

d. Pemilihan Fakta

Penentuan data dalam tabel fakta pada proses perancangan data *warehouse* di Perusahaan Penyedia Layanan Perjalanan PT. Bahana Sukses Sejahtera. Tabel fakta dimanfaatkan untuk menghubungkan informasi dari dimensi data yang saling terhubung. Tabel-tabel dari fakta tersebut di gambarkan dalam diagram seperti yang tercantum pada gambar 3 berikut ini:



Gambar 3. Fakta Jemaah Umroh

e. Penyimpanan *Pre-Calculation* di Tabel Fakta
 Pengkajian ulang dapat dilakukan setelah memilih fakta untuk menentukan apakah proses *precalculation* dapat dilakukan pada tahap berikutnya.

f. Memastikan Tabel Dimensi
 Penggambaran secara detail untuk menghasilkan keterangan yang lebih lengkap untuk dimensi data untuk memudahkan pengguna dalam memahami penggambaran data dalam bentuk visual.

g. Pemilihan Durasi *Database*
 Perkiraan data dilakukan dengan memilih rangkaian data dalam 3 periode tahun sebelum masa pandemic Covid-19 merebak. Periode tahun tersebut dilakukan pada tahun

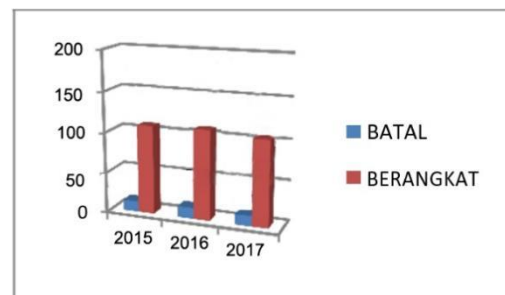
2015, 2016 dan tahun 2017. Periode tahun-tahun tersebut dapat memberikan gambaran untuk memprediksi keadaan informasi yang dicari yaitu tentang Jemaah umroh yang tidak jadi berangkat melalui perusahaan PT. Bahana Sukses Sejahtera.

- h. Pelacakan data dari perubahan perlahan dari dimensi data. Atribut yang tersedia pada tabel dimensi tidak selamanya bernilai tetap sehingga suatu saat data tersebut pada suatu saat nilainya berubah.
- i. Menentukan Prioritas dan *Query* yang tepat.
 Data warehouse dirancang sesuai dengan kapasitas data karena akan berpengaruh dengan daya tamping media penyimpanan data. Perlu diperhatikan pada saat merancang apakah proses perancangan data membuat kapasitas data semakin besar. Sebaiknya disiapkan media penyimpanan data untuk kapasitas data yang besar.

2.5 Presentasi Data *Warehouse*

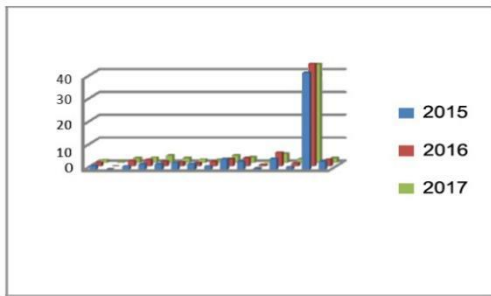
Berdasarkan data yang tersimpan didalam database sistem informasi perusahaan PT. Bahana Sukses Sejahtera selama periode tahun 2015, 2016 dan 2017 yang akan digunakan sebagai data training.

- a. Berdasarkan Dimensi Waktu
 Pengukuran jumlah Jemaah umroh yang digambarkan dalam dimensi data berdasarkan periode waktu tertentu. Contohnya adalah jumlah jemaah umroh selama tahun 2015 adalah berjumlah 327 orang.
- b. Berdasarkan Dimensi Jemaah
 Penggambaran dimensi data jemaah berupa tabel data jemaah umroh yang rentan terhadap pembatalan keberangkatan karena memiliki riwayat penyakit sehingga memberikan informasi bisa dan tidaknya jemaah untuk berangkat.



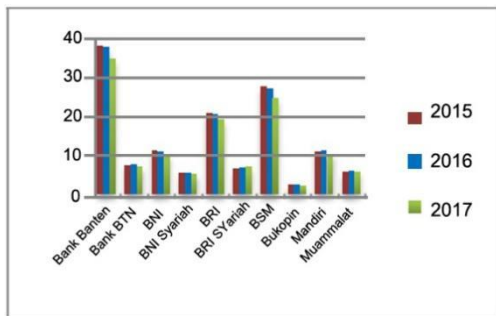
Gambar 6 Grafik Dimensi Jemaah dengan Kategori Keterangan

c. Berdasarkan Dimensi Wilayah
 Penggambaran data berdasarkan wilayah tempat tinggal Jemaah umroh untuk membuat dimensi wilayah berdasarkan daerah Kabupaten dan Kota madia sekitar Prosinsi Banten. Contohnya adalah jumlah jemaah umroh dari Kota Tangerang pada tahun 2017 sebanyak 43 orang.



Gambar 4. Grafik Dimensi Wilayah

- d. Berdasarkan Dimensi Bank Perancangan dengan data dimensi bank tempat rekening Jemaah umroh disimpan menghasilkan tabel bank. Berdasarkan database sistem informasi perusahaan ada sekitar Tahun 2017 ada 38 jiwa jemaah umroh membuka rekening untuk biaya perjalanan umroh di Bank Pembangunan Banten.



Gambar 5. Grafik Dimensi Bank

2.6 Penerapan Data Mining

Proses dalam tahapan pemanfaatan pengolahan data mining adalah sebagai berikut:

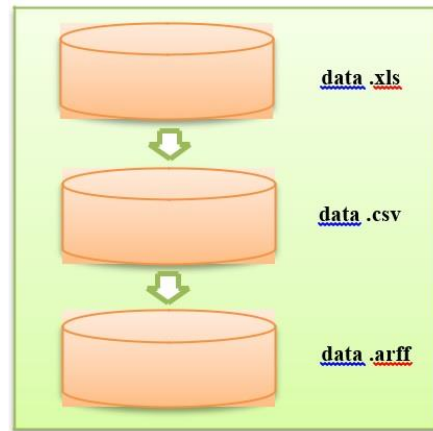
- a. Pengumpulan Data

Tahapan pengumpulan data dilakukan sesuai dengan kebutuhan dalam penerapan data mining. Penggunaan data-data dalam periode selama tiga tahun data training yaitu 2015, 2016 dan 2017 dan sebagai data *testing*.
- b. Pra Proses Data atau *Cleansing* Data

Pembersihan data dilakukan untuk menghapus data yang tidak lengkap dan data yang tidak diperlukan pada proses data mining. Data yang terpilih merupakan data yang menjadi fokus dalam penelitian.
- c. Transformasi Data

Perubahan bentuk data dalam proses transformasi merubah dari bentuk data satu ke bentuk data lain. Transformasi data tersebut disesuaikan dengan perangkat lunak yang digunakan dalam proses data mining. Perangkat yang digunakan adalah Aplikasi bernama WEKA dimana perangkat ini hanya dapat membaca data dengan format file *.adff* dan *.csv*. Master data dengan format file lain dapat dirubah terlebih dahulu menjadi format yang

sesuai seperti *.csv*. Berikut merupakan transformasi data dari bentuk data excel menjadi format csv yang bisa dibaca oleh aplikasi WEKA.



Gambar 7. Transformasi Data

- d. Penerapan menggunakan aplikasi WEKA

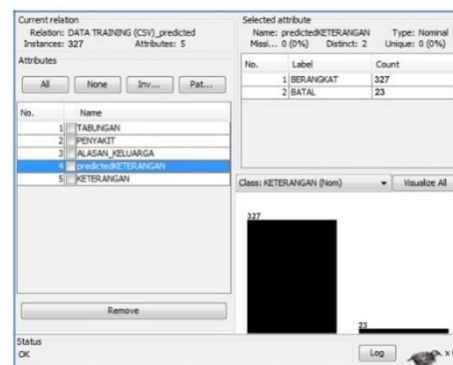
Perangkat lunak WEKA yang digunakan dalam penelitian ini digunakan untuk membuat proses prediksi. Proses penerapan menggunakan perangkat lunak WEKA untuk mendapatkan informasi yang dicari adalah sebagai berikut:

 - Memasukan data training
 - Penggunaan Algoritma *Nearest Neighbor*
 - Memasukan data testing
 - Informasi pengetahuan berupa prediksi
- e. Informasi Hasil Analisis

Proses analisa data menggunakan perangkat lunak WEKA menghasilkan berbagai informasi dan ilmu pengetahuan sebagai berikut:

 - i. Hasil proses perhitungan dan pengolahan data Didapatkan hasil prediksi meliputi 327 calon jemaah umroh berangkat sedangkan 23 jemaah umroh batal berangkat.
 - ii. Prediksi hasil perhitungan menggunakan perangkat lunak WEKA yang menjadi factor dominan penyebab keberangkatan calon Jemaah umroh batal untuk berangkat.

Penggambaran data hasil prediksi tersebut dengan menggunakan perangkat lunak WEKA dapat terlihat seperti pada tampilan gambar berikut:



Gambar 8. Hasil Prediksi Data Testing

Gambar diatas dapat menghasilkan informasi berupa prediksi Jemaah umroh yang berangkat dan Jemaah umroh membatalkan keberangkatan dengan alasan tertentu.

4. KESIMPULAN

Berdasarkan proses pengolahan data-data dari perusahaan penyedia jasa layanan perjalanan umroh PT. Bahana Sukses Sejahtera menghasilkan informasi-informasi yang dapat dijadikan sebagai ilmu pengetahuan dalam memberikan masukan dalam pengambilan keputusan perusahaan. Berikut ini adalah beberapa kesimpulan yang dihasilkan dalam pengembangan penelitian ini:

Periode tahun 2015, 2016 dan tahun 2017 penyakit stroke menjadi faktor dominan yang mempengaruhi pembatalan keberangkatan jemaah umroh. Alasan keluarga menjadi faktor kedua yang menjadi faktor pembatalan keberangkatan.

Jemaah umroh pada perusahaan PT. Bahana Sukses Sejahtera banyak mengandalkan Bank Pembangunan Banten merupakan Bank yang menjadi pilihan paling banyak jemaah membuka rekeningnya di sana. Tahun 2015 memiliki 136 nasabah jemaah umroh, periode tahun 2016 mempunyai 115 nasabah jemaah umroh dan periode tahun 2017 berjumlah 97 nasabah jemaah umroh. Penyebaran asal daerah para jemaah umroh paling banyak dari Kota Tangerang selama 3 tahun. Jemaah yang paling sedikit juga dari Kota Tangerang. Daerah dengan jemaah paling sedikit berangkat adalah dari daerah Serang.

Hasil pengolahan data berdasarkan data testing dengan data yang diujikan sebanyak 348 data dengan hasil prediksi berjumlah 288 calon jemaah umroh dapat berangkat dan 39 calon jemaah umroh membatalkan diri karena berbagai alasan.

Tabungan jemaah menjadi faktor menjadi dominan selain alasan keluarga menjadi alasan jemaah tidak jadi berangkat. Faktor berikutnya adalah penyakit menjadi alasan jemaah membatalkan keberangkatan dimana beberapa penyakit seperti Diabetes dan tekanan darah tinggi menjadi penyakit paling banyak diderita calon jemaah umroh batal berangkat.

Tabungan di bank jemaah umroh menjadi kondisi paling tidak memungkinkan untuk berangkat yang tidak cukup di tambah dengan kondisi kesehatan yang kurang baik serta alasan faktor keluarga. Faktor kompleks tersebut dapat diprediksi berdasarkan analisis *nearest neighbor* dan algoritma dapat dipastikan calon jemaah tidak bisa berangkat

DAFTAR PUSTAKA

[1] Connolly, Thomas, C. B 2010, Database System: A Practical Approach to Design, Implementation, and Management Fifth Edition, Pearson Education Inc.

- [2] Sulanta, Feri 2010, Data Mining: Meramalkan Bisnis Perusahaan, Elex Media Komputindo, Jakarta.
- [3] Ponniah, Paulraj 2001, Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals, A Wiley-Interscience Publication, New York.
- [4] Han, J, Kamber, M & Pei, J 2011, Data Mining: Concepts and Techniques Third Edition, Elsevier, United States Of America.
- [5] Larose, Daniel T. 2005. Discovering Knowledge in Data: An Introduction to Data Mining. John Willey & Sons, Inc.
- [6] Kusriani 2009, Algoritma Data Mining, Andi Offset, Yogyakarta.
- [7] Inmon, W, H 2005, Building The Data Warehouse Fourth Edition, Wiley Publishing, Inc., United States Of America.
- [8] Kimball, R & Ross, M 2013, The Data Warehouse Toolkit Third Edition, John Wiley and Sons, Canada.
- [9] Wirama, K, Sudianto, H & Hermawan, Y 2009, The Essential Business Intelligence in Microsoft SQL Server 2008, Indc, Jakarta.