
Penentuan Produk Asuransi Bpjs Berdasarkan Profil Pelanggan Dengan Pendekatan K-Nearest Neighbor Manhattan Distance

Determination BPJS Insurance Products Based on Customer Profile With K-Nearest Neighbor Approach Manhattan Distance

Titis Dwi Rahmawati, Fajrian Nur Adnan, M.CS

Jurusan Sistem Informasi, Ilmu Komputer, Universitas Dian Nuswantoro

Fakultas Ilmu Komputer Universitas Dian Nuswantoro Semarang

Jalan Nakula I No. 5-11 Semarang 50131

Telp : (024) 3517261, Fax : (024) 3520165

E-mail : 112201204714@mhs.dinus.ac.id¹, fajrian@dsn.dinus.ac.id²

Abstrak

Dengan semakin berkembangnya kebutuhan masyarakat asuransi menjadi hal yang diperlukan dan sangat dibutuhkan oleh masyarakat. Dalam kasus asuransi jiwa produk yang ditawarkan oleh perusahaan asuransi sangat beragam. Hal ini menyebabkan *customer* mengalami kebingungan dalam menentukan pilihan produk asuransi yang akan mereka beli. Padahal setiap produk memiliki fungsi yang berbeda-beda. *Customer* harus memilih produk tersebut dengan teliti sesuai dengan kebutuhan sehingga tidak salah memilih produk dan mengalami kerugian dengan membeli produk asuransi yang salah. Laporan tugas akhir ini akan membahas tentang memprediksi produk yang tepat untuk dipromosikan terhadap *customer* sesuai dengan *profile customer*. Dengan menggunakan data pelanggan yang diperoleh dari BPJS Cabang II Semarang yang ada akan dilakukan perhitungan yang akan menghasilkan prediksi produk asuransi yang akan dipromosikan terhadap pelanggan, menggunakan algoritma *K-Nearest Neighbor Manhattan Distance*, serta penggunaan MATLAB untuk pengotomatisasian perhitungan *KNN Manhattan Distance*. Dari laporan ini dihasilkan program yang diharapkan dapat memberikan rekomendasi produk untuk ditawarkan terhadap *customer* sesuai dengan *profile customer*.

Kata kunci : Algoritma KNN, Data Mining, Manhattan Distance, Asuransi, Klasifikasi, Penentuan Produk Asuransi

Abstract

Nowdays the needs in society is increasing and so diverse, this make insurance becomes necessary and needed by society. There is so many variant of life insurance product, this make the customer confused to choose which life insurance product they should buy. In fact, product have different function. Customers should choose the products carefully according to the their need so that no one choose the wrong insurance products and experience loss. This research will discuss about the right product to be promoted to the customer in accordance with the customer profile. Using customer data obtained from BPJS Cabang II Semarang predictions of insurance products to be promoted to customers will be calculated using K-Nearest Neighbor Algorithym with Manhattan Distance, as well as the use of MATLAB for automating the calculation of KNN Manhattan Distance. The outcome of this research is a MATLAB program that is expected to provide product recommendation who can be offered to the cutomer in accordance with the customer profile.

Key words : KNN Algorithym, Data Mining, Manhattan Distance, Insurance, Classification, Insurance Product Recommendation

1. PENDAHULUAN

Asuransi telah menjadi sebuah kebutuhan yang lazim di era yang semakin berkembang dan maju seperti saat ini. Dengan semakin berkembangnya kebutuhan masyarakat, asuransi menjadi hal yang diperlukan dan sangat dibutuhkan. Perusahaan asuransi menawarkan banyak produk asuransi seperti asuransi jiwa, asuransi pendidikan, asuransi kendaraan, asuransi haji. Semakin berkembangnya produk yang ditawarkan oleh perusahaan ini didorong oleh kemajuan industri asuransi yang semakin berkembang di Indonesia [1].

Persaingan antar perusahaan untuk meningkatkan volume penjualan produk semakin ketat oleh sebab itu perusahaan perlu melakukan hal hal inovatif agar mereka mampu bersaing dengan perusahaan asuransi yang lain. Dalam kasus asuransi, produk yang ditawarkan oleh perusahaan asuransi sangat beragam. Hal ini menyebabkan *customer* mengalami kebingungan dalam menentukan pilihan produk asuransi yang akan dibeli. Padahal setiap produk memiliki fungsi yang berbeda-beda. *Customer* harus memilih produk tersebut dengan teliti sesuai dengan kebutuhan sehingga tidak salah memilih produk dan mengalami kerugian dengan membeli produk asuransi yang salah.

BPJS Ketenagakerjaan merupakan salah satu perusahaan asuransi yang ada di Indonesia yang melayani masyarakat dengan jasa asuransi ketenagakerjaan. Ada 4 produk yang ditawarkan oleh BPJS Ketenagakerjaan, yaitu JKK (Jaminan Kecelakaan Kerja), JHT (Jaminan Hari Tua), JKM (Jaminan Kematian) dan JP (Jaminan Pensiun). Berdasarkan data yang diperoleh dari pihak BPJS jumlah pengguna asuransi ketenagakerjaan cukup banyak. Dari jumlah tersebut dapat diketahui bahwa pengguna asuransi ketenagakerjaan sangatlah beragam, mereka memiliki rentang umur, pekerjaan, dan penghasilan yang berbeda beda. Dengan kriteria pengguna yang berbeda-beda perusahaan harus mampu mengetahui kebutuhan asuransi jenis apakah yang paling tepat untuk pengguna tersebut.

Perusahaan memiliki data yang cukup besar. Data tersebut apabila diolah menggunakan teknologi dengan benar maka akan menghasilkan informasi yang dapat membantu dalam strategi kemajuan perusahaan. Dengan manfaat yang dapat diperoleh dari pengolahan data tersebut maka diperlukan teknologi dan metode yang tepat untuk mengolah data tersebut agar dapat menghasilkan informasi yang bermanfaat bagi perusahaan.

Mengolah data tersebut guna memprediksi minat konsumen terhadap asuransi sangatlah membantu perusahaan guna memberikan rekomendasi jenis produk asuransi yang tepat bagi seorang konsumen. Metode *classification* merupakan salah satu metode yang digunakan dalam mengolah data. *Classification* merupakan proses guna menentukan model maupun fungsi yang menjelaskan atau membedakan konsep atau kelas data yang bertujuan untuk memperkirakan kelas dari sebuah objek yang labelnya tidak diketahui. Proses *classification* membentuk model yang dapat mengelompokkan data kedalam kelas-kelas tertentu berdasarkan aturan dan fungsi tertentu. Model tersebut dapat berupa pohon keputusan ataupun formula matematis [2].

Berdasarkan *International Conference on Data Mining (ICDM)* menyatakan bahwa algoritma KNN merupakan salah satu algoritma terbaik dalam *Top 10 algorithm in data mining* [3]. Beberapa penelitian menggunakan algoritma KNN sudah dilakukan dan memberikan hasil yang baik. Selain sebagai salah satu algoritma klasifikasi tertua dan algoritma yang mudah digunakan, KNN memberikan solusi klasifikasi berdasarkan dari kelas terdekat dari k parameter.

Mengingat manfaat dari pengolahan data yang dapat menghasilkan informasi yang bermanfaat bagi perusahaan maka penulis bermaksud untuk melakukan penelitian dengan mengolah data sebuah perusahaan asuransi guna memprediksi produk asuransi yang tepat untuk dipromosikan terhadap pelanggan yang sesuai dengan kriteria pelanggan.

2. METODE PENELITIAN

Objek penelitian pada tugas akhir ini adalah BPJS Cabang II Semarang, perusahaan ini akan dibuatkan program untuk memprediksi produk asuransi yang akan dipromosikan terhadap konsumen menggunakan pendekatan K-Nearest Neighbor Manhattan Distance.

Algoritma KNN termasuk dalam jenis klasifikasi pada data mining. Algoritma *k-nearest neighbor* (k-NN atau KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. KNN bertujuan untuk menemukan pola baru dalam data dengan menghubungkan pola data yang sudah ada dengan data yang baru. Contoh aturan asosiasi dalam sebuah perusahaan asuransi yaitu dapat mengetahui pola dari pelanggan dalam menentukan produk asuransi. Dengan pengetahuan tersebut maka perusahaan dapat melakukan prediksi produk asuransi yang akan dipromosikan terhadap pelanggan baru.

Tujuan dari algoritma *k-NN* adalah untuk mengklasifikasi objek baru berdasarkan atribut dan *training samples*. Dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada *k-NN*. Pada proses pengklasifikasian, algoritma ini tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Algoritma *k-NN* menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru.

Algoritma *k-NN* adalah algoritma yang menentukan nilai jarak pada pengujian *data testing* dengan *data training* berdasarkan nilai terkecil dari nilai ketetanggaan terdekat didefinisikan sebagai berikut:

$$D_{nn}(C_1, C_2) = \min_{1 \leq i \leq r, 1 \leq j \leq s} d(y_i, z_j)$$

Langkah-langkah KNN:

1. Menentukan parameter K (jumlah tetangga paling dekat).
2. Menghitung kuadrat jarak Euclid (query instance) masing-masing objek terhadap data sampel yang diberikan.
3. Kemudian mengurutkan objek-objek tersebut kedalam kelompok yang mempunyai jarak Euclid terkecil.
4. Mengumpulkan kategori Y (Klasifikasi nearest neighbor).
5. Dengan menggunakan kategori nearest neighbor yang paling mayoritas maka dapat diprediksikan nilai query instance yang telah dihitung.

3. HASIL DAN PEMBAHASAN

3.1 Data Pelanggan

Data pelanggan yang diperoleh dari BPJS Ketenagakerjaan Cabang II Semarang berupa data primer dalam bentuk excel. Data tersebut berisikan field-field yang menggambarkan ciri-ciri seorang customer yang terdiri dari no, nama, sex, alamat, lahir, perusahaan, upah dan program. Data tersebut perlu diolah terlebih dahulu kedalam dataset SQLyog sebelum nantinya diolah dalam program perhitungan MATLAB untuk melakukan perhitungan.

NO	NAMA	SEX	ALAMAT	LAHIR	PERUSAHAAN	UPAH	PROGRAM
1	KUSNADI	L	JALI RT 01 RW 01 BONANG	18-02-1976	SUBUR MAKMUR MIGAS PRATAMA PT	1,909,000	JKK, JKM
2	SUROTO	L	PURWOREJO RT03 RW01 BONANG DEMAK	28-04-1988	MAKMUR JAYA MIGAS PRATAMA PT	1,941,000	JKK, JKM
3	SUPRAPTO	L	TEGALREJO RT 02 RW 09 KEL KEMIJEN KEC SEMARANG TIMUR KAB SEMARANG	28-07-1984	QUARTINDO SEJATI FURNITAMA PT	1,941,000	JKK, JHT
4	RINI NIRWATI	P	JOGO IV RT05 RW02 KEC SAYUNG	09-11-1988	ALTEK PRECISION MACHINERY PT	1,941,000	JKK, JKM
5	SITI MUNIFAH	P	DK KRAJAN LOR RT 001/003 JEPANGPAKIS JATI KUDUS	31-05-1981	LASINDO METAL UTAMA CV	1,941,000	JKK, JKM
6	FEBRI SULISTYANA	P	DS WONOLOPO RT04 RW01 KEC MIJEN KAB SEMARANG	12-02-1989	UNIVERSAL JIROLU CORPORATION CV	1,976,407	JKM
7	SUMARDIYANI	L	DS SARI RT02 RW04 KEL SARI KEC GAJAH KAB DEMAK	16-07-1988	UPPKH KECAMATAN KRADENAN	1,981,000	JKK, JKM
8	ZENI AGGRAENI	P	TANGGUNG HARJO RT03 RW04 KELBALEROMO KEC DEMPET KAB DEMAK	01-07-1977	UPPKH KECAMATAN GABUS	1,981,000	JKK, JKM
9	MASRUM	L	BUARAN RT 09 RW 01 MAYONG JEPARA	29-11-1989	UPPKH KECAMATAN BRATI	2,013,000	JKK, JKM
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
1499	DYAH DEWI HARSANTI	P	NGELOWETAN, 001/001, MIJEN, DEMAK	29-08-1984	VICTORY APPAREL TKA PT	1,909,000	JKK, JHT, JKM

Gambar 1 Data Pelanggan

3.2 Pre-processing data

Step pertama yang dilakukan dalam *pre-processing* data yaitu penyeleksian data atau *data selection*, namun karena data yang diperoleh dari BPJS Ketenagakerjaan Cabang Semarang hanya ada satu data yaitu data konsumen maka tidak perlu dilakukan proses data selection. Karena tidak perlu dilakukannya proses data selection maka proses dilanjutkan pada tahap data cleaning.

Step yang kedua yang dilakukan yaitu *pre-processing / cleaning*, pada *pre-processing / cleaning* dilakukan proses penghilangan attribute-attribute yang tidak diperlukan dalam proses penghitungan supaya tidak memakan banyak *space*. Data konsumen perlu dihilangkan dari kesalahan data yang ada seperti data yang sama, inkonsisten data dan kesalahan data. Data konsumen yang semula terdiri dari no urut, nama, sex, alamat, lahir, perusahaan, gaji dan program, dalam tahapan ini data tersebut akan dijadikan menjadi no, sex, lahir, gaji dan program. Dalam data konsumen alamat dan perusahaan perlu dihilangkan hal ini dikarenakan alamat dan perusahaan tidak digunakan dalam proses data mining karena memiliki keinkonsistenan data. Untuk mendapatkan tabel tersebut dilakukan proses query sebagai berikut

select no, sex, lahir, gaji, program from bpjs

berikut hasil dari *pre-processing / cleaning*.

NO	NAMA	SEX	ALAMAT	LAHIR	PERUSAHAAN	UPAH	PROGRAM
1	KUSNADI	L	JALI RT 01 RW 01 BONANG	18-02-1976	SUBUR MAKMUR MIGAS PRATAMA PT	1,909,000	JKK, JKM
2	SUROTO	L	PURWOREJO RT03 RW01 BONANG DEMAK	28-04-1988	MAKMUR JAYA MIGAS PRATAMA PT	1,941,000	JKK, JKM
3	SUPRAPTO	L	TEGALREJO RT 02 RW 09 KEL KEMIJEN KEC SEMARANG TIMUR KAB SEMARANG	28-07-1984	QUARTINDO SEJATI FURNITAMA PT	1,941,000	JKK, JHT
4	RINI NIRWATI	P	JOGO IV RT05 RW02 KEC SAYUNG	09-11-1988	ALTEK PRECISION MACHINERY PT	1,941,000	JKK, JKM
5	SITI MUNIFAH	P	DK KRAJAN LOR RT 001/003 JEPANGPAKIS JATI KUDUS	31-05-1981	LASINDO METAL UTAMA CV	1,941,000	JKK, JKM

↓

NO	SEX	LAHIR	UPAH	PROGRAM
1	'L'	1976	1909000	JKK,JKM'
2	'L'	1988	1941000	JKK,JKM'
3	'L'	1984	1941000	JKK,JHT'
4	'P'	1988	1941000	JKK,JKM'
5	'P'	1981	1941000	JKK,JKM'

Gambar 2 Perubahan Data pada Proses Pre-processing / Cleaning

Setelah attribute selection tahap selanjutnya adalah transformation, dimana pengimplementasian algoritma ke dalam data yang telah dipilih sehingga data tersebut sesuai untuk proses data mining. Langkah awal yang harus dilakukan dalam proses ini adalah mengkoneksikan data konsumen ke dalam MATLAB. Data tersebut perlu dikoneksikan agar menjadi database. Proses ini bertujuan agar nantinya data berupa satu file dan tidak diperlukan banyak data untuk diproses. Setelah data tersebut diimpor menjadi database maka data tersebut dikoneksikan dengan MATLAB untuk kemudian dilakukan proses selanjutnya yaitu *data transformation*. Dalam tahapan ini terdapat proses dimana variable sex dan program yang semula berupa character akan dirubah sesuai dengan aturan ascii.

Variable sex terdiri dari L dan P maka akan dirubah kedalam bilangan decimal sesuai dengan aturan ASCII. L akan berubah menjadi 76 sedangkan P akan berubah menjadi 80. Variable Program yang berisi jenis program yang diambil juga akan dikonversi kedalam bilangan decimal sesuai aturan ASCII. Ada 4 jenis program yang ada dalam BPJS yaitu JHT, JKK, JKM dan JP yang apabila dikonversikan dalam ASCII menjadi 230, 224, 226, 154. Berikut merupakan contoh perhitungan variable Program yang dibahas dalam tabel di bawah ini.

JKK, JKM	224+44+32+226+13=539
JKK, JHT	224+44+32+230+13=543
JKM	226+13=239
JHT, JKM	230+44+32+226+13=545
JKK, JHT, JKM	224+44+32+230+44+32+226+13=845
JKK, JHT, JKM,JP	224+44+32+230+44+32+226+44+154+13=1043
JKK, JKM,JP	224+44+32+226+44+154+13=737
JKK, JHT, JP	224+44+32+230+44+154+13=773
JHT, JKM,JP	230+44+32+226+44+154+13=743

Gambar 3 Tahap Konversi ASCII

Hal ini dilakukan guna memudahkan dalam proses pengimplementasian algoritma. Hasil dari proses perubahan tersebut dapat dilihat pada gambar di bawah ini

NO	SEX	LAHIR	UPAH	PROGRAM
1	'L'	1976	1909000	'JKK, JKM
2	'L'	1988	1941000	'JKK, JKM
3	'L'	1984	1941000	'JKK, JHT
4	'P'	1988	1941000	'JKK, JKM
5	'P'	1981	1941000	'JKK, JKM
6	'P'	1989	1976407	'JKM
7	'L'	1988	1981000	'JKK, JKM
8	'P'	1977	1981000	'JKK, JKM
9	'L'	1989	2013000	'JKK, JKM
10	'L'	1989	2013000	'JKK, JHT

↓

NO	SEX	LAHIR	UPAH	PROGRAM
1	76	1976	1909000	539
2	76	1988	1941000	539
3	76	1984	1941000	543
4	80	1988	1941000	539
5	80	1981	1941000	539
6	80	1989	1976407	239
7	76	1988	1981000	539
8	80	1977	1981000	539
9	76	1989	2013000	539
10	76	1989	2013000	543

Gambar 4 Perubahan Variabel Sex dan Program

3.2 Proses Perhitungan dan hasil program

Setelah dilakukannya *pre-processing* data step yang dilakukan selanjutnya yaitu proses peritungan, pada proses perhitungan di kasus ini memerlukan extra tenaga karena data yang dihitung sangatlah banyak, transaksi yang digunakan untuk proses perhitungan ini adalah 1499 record, jadi supaya proses perhitungan yang dilakukan tidak memakan waktu lama dan menyusahkan, disini dibuatlah program perhitungan menggunakan algoritma *K-Nearest Neighbor* untuk menemukan pola pelanggan dalam menentukan produk asuransi sehingga dapat dilakukan untuk memprediksi produk yang akan dipromosikan kepada pelanggan. Berikut merupakan tampilan program perhitungan yang dibuat



Gambar 5 Screenshoot Program

Dalam program tersebut yang perlu dilakukan hanya menentukan sex, lahir dan upah dari kriteria pelanggan lalu setelah itu menekan tombol check maka akan muncul hasil prediksi produk asuransi yang dapat dipromosikan terhadap pelanggan berdasarkan kriteria yang telah diinputkan. Untuk membuktikan bahwa proses perhitungan ini benar maka dilakukan proses pengujian dengan cara perhitungan secara manual seperti berikut.

Dalam program *K-Nearest Neighbor* yang dibuat telah memuat 1499 data pelanggan yang akan digunakan untuk melakukan customer profiling. Dalam perhitungan ini akan dipilih sejumlah data untuk dijadikan sample dalam proses perhitungan manual ini. Dari 1499 data pelanggan yang telah diperoleh akan dipilih 50 data random untuk dijadikan sample dalam melakukan perhitungan manual, berikut merupakan proses perhitungan manual tersebut. Berikut merupakan 50 data random yang dijadikan sample tersebut.

NO	SEX	LALIR	UPAH	PROGRAM
482	76	1979	2088000	539
796	76	1987	2088000	1043
692	76	1983	2080000	845
850	80	1985	2106000	743
513	76	1991	2038000	845
856	76	1986	2113000	743
466	80	1989	2038000	539
265	80	1990	2038000	545
630	76	1971	2048000	539
563	76	1982	2044000	1043
542	76	1979	2038000	545
394	80	1982	2038000	543
421	76	1991	2038000	845
138	76	1985	2018000	545
293	76	1987	2038000	543
8	80	1977	1981000	539
933	76	1975	2128000	1043
610	76	1988	2044000	1043
892	76	1986	2123500	845
421	76	1991	2038000	845
401	80	1985	2038000	545
554	76	1988	2038000	845
597	80	1988	2044000	845
507	76	1986	2038000	539
588	80	1990	2044000	539
133	76	1989	2018000	545
762	76	1985	2086000	1043
652	76	1992	2068000	473
352	80	1983	2038000	543
313	80	1990	2038000	545
275	76	1989	2038000	845
150	80	1989	2028000	845
335	80	1986	2038000	543
834	80	1979	2093000	1043
234	76	1986	2038000	545
70	76	1992	2013000	545
495	76	1992	2038000	845
123	76	1988	2018000	845
403	80	1987	2038000	545
92	76	1982	2013000	539
721	76	1988	2081000	845
343	80	1990	2038000	545
567	76	1980	2044000	1043
655	76	1987	2068000	545
653	76	1988	2068000	539
701	76	1987	2080000	845
302	80	1984	2038000	545
86	76	1984	2013000	845
634	76	1992	2053000	539
267	76	1987	2038000	845

Gambar 6 Data Sampel

Tahapan awal dari proses perhitungan manual algoritma K-Nearest Neighbor adalah menentukan nilai k, dan nilai k yang akan digunakan tersebut adalah 7. Nilai k ditentukan dengan bilangan ganjil dikarenakan nilai k yang ganjil untuk jumlah kelas yang genap akan memudahkan voting karena akan mengurangi kemungkinan terjadinya dua kelas yang mendapat voting yang sama. Setelah ditentukan jumlah k maka akan dihitung jarak setiap data training tersebut terhadap setiap variable [14].

Berikut merupakan contoh menghitung jarak terhadap data training yang dilakukan menggunakan metode perhitungan jarak *Manhattan Distance* atau *city block distance*. Dengan menjadikan data ke 796 sebagai data uji dalam perhitungan ini.

1. $D(796,1) = |76 - 76| + |1987 - 1987| + |2088000 - 1909000| = 179011$
2. $D(796,2) = |76 - 76| + |1987 - 1988| + |2088000 - 1941000| = 147001$
3. $D(796,3) = |76 - 76| + |1987 - 1984| + |2088000 - 1941000| = 147003$
4. $D(796,4) = |76 - 80| + |1987 - 1988| + |2088000 - 1941000| = 147005$
5. $D(796,5) = |76 - 76| + |1987 - 1981| + |2088000 - 1941000| = 14010$

Dengan melakukan perhitungan yang sama seperti contoh diatas maka didapatkan hasil perhitungan jarak terhadap data ke 796 seperti pada tabel berikut ini.

Data ke	Sex	Lahir	Upah	Program	Distance
D(796,1)	76	1976	1909000	539	179011
D(796,2)	76	1988	1941000	539	147001
D(796,3)	76	1984	1941000	543	147003
D(796,4)	80	1988	1941000	539	147005
D(796,5)	80	1981	1941000	539	147010
D(796,6)	80	1989	1976407	239	111599
D(796,7)	76	1988	1981000	539	107001
D(796,9)	76	1989	2013000	539	75002
D(796,10)	76	1989	2013000	543	75002
.
.
.
D(796,1499)	80	1984	1909000	845	179007

Gambar 7 Tabel Hasil Perhitungan Jarak Data Sampel 796

Setelah dilakukan perhitungan jarak langkah selanjutnya adalah mengurutkan jarak tersebut berdasarkan kedekatannya. Berikut ini adalah hasil pengurutan jarak tersebut.

Data ke	Sex	Lahir	Upah	Program	Distance	Rank
1	76	1976	1909000	539	179011	1342
2	76	1988	1941000	539	147001	1264
3	76	1984	1941000	543	147003	1265
4	80	1988	1941000	539	147005	1266
5	80	1981	1941000	539	147010	1267
6	80	1989	1976407	239	111599	1170
7	76	1988	1981000	539	107001	1169
9	76	1989	2013000	539	75002	826
10	76	1989	2013000	543	75002	827
11	80	1987	2013000	539	75004	860
12	80	1982	2013000	545	75009	901
13	80	1989	2013000	545	75006	883
14	76	1989	2013000	539	75002	828
15	76	1978	2013000	539	75009	902
16	76	1988	2013000	545	75001	810
17	76	1973	2013000	239	75014	909
18	76	1987	2013000	539	75000	803
19	80	1977	2013000	539	75014	910
.
.
739	76	1987	2088000	1043	0	1
.
.
1450	80	1984	1909000	845	179007	1338

Gambar 8 Hasil Perankingan Jarak

Dari hasil perankingan tersebut perlu dipilih jarak terdekat berdasarkan nilai k yang telah ditentukan, dalam hal ini akan dilakukan perhitungan dengan menggunakan nilai k 1, 3, 5,7, 11 maka akan dipilih ranking sesuai dengan nilai k data yang memiliki nilai kedekatan paling besar. Hasilnya ada dalam tabel berikut ini.

Data ke	Sex	Lahir	Uyah	Program	Distance	Rank
739	76	1987	2088000	1043	0	1
786	76	1987	2088000	743	0	2
799	76	1987	2088000	1043	0	3
810	76	1987	2088000	1043	0	4
741	76	1988	2088000	845	1	5
748	76	1986	2088000	539	1	6
750	76	1988	2088000	545	1	7

Gambar 9 Hasil Perangkingan Jarak Sesuai Dengan Nilai K 7

Berdasarkan tabel di atas dapat dilihat bahwa dengan nilai $k=7$ maka data dengan jarak terdekat terdiri dari data ke 739, 799, 810 dengan kelas program 1043, data ke 786 dengan kelas program 743, data ke 741 dengan kelas program 845, data ke 748 dengan kelas program 539 dan data ke 750 dengan kelas program 545. Oleh karena itu maka dengan nilai $k=7$ dapat disimpulkan bahwa data uji diprediksi masuk dalam kelas program 1034, hal ini dikarenakan jumlah data dengan kelas program 1043 lebih banyak yaitu muncul 3 kali. Kelas program 1043 apabila dikonversikan merupakan program JKK, JKM, JHT dan JP.

Maka dapat disimpulkan bahwa hasil dari program dan perhitungan manual memiliki hasil yang sama, hal ini dapat dilihat dari gambar di bawah ini.



Gambar 10 Perbandingan Hasil Program dan Hasil Perhitungan Manual

3.3 Evaluasi dan Pembahasan

Pola informasi yang dihasilkan dari proses data mining, perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dan proses KDD yang disebut interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya. Dibuktikan dari perhitungan yang dihitung sebelumnya, maka proses evaluation akan dibuktikan dengan data perbandingan hasil berikut ini.

Data ke	Fakta	Hasil Program
482	'JKK,JKM'	'JHT,JKM'
796	'JKK,JHT,JKMLP'	'JKK,JHT,JKMLP'
692	'JKK,JHT,JKM'	'JKK,JHT,JKM'
850	'JHT,JKMLP'	'JHT,JKMLP'
513	'JKK,JHT,JKM'	'JKK,JHT,JKM'
856	'JHT,JKMLP'	'JHT,JKMLP'
466	'JKK,JKM'	'JKK,JHT,JKM'
265	'JHT,JKM'	'JKK,JHT,JKM'
630	'JKK,JKM'	'JKK,JKM'
563	'JKK,JHT,JKMLP'	'JKK,JHT,JKM'
542	'JHT,JKM'	'JHT,JKM'
394	'JKK,JHT'	'JKK,JHT'
421	'JKK,JHT,JKM'	'JKK,JHT,JKM'
138	'JHT,JKM'	'JKK,JHT'
293	'JKK,JHT'	'JKK,JHT,JKM'
8	'JKK,JKM'	'JKK,JKM'
933	'JKK,JHT,JKMLP'	'JKK,JHT,JKMLP'
610	'JKK,JHT,JKMLP'	'JKK,JHT,JKMLP'
892	'JKK,JHT,JKM'	'JKK,JHT,JKMLP'
421	'JKK,JHT,JKM'	'JKK,JHT,JKM'
401	'JHT,JKM'	'JHT,JKM'
554	'JKK,JHT,JKM'	'JKK,JHT,JKM'
597	'JKK,JHT,JKM'	'JKK,JHT,JKM'
507	'JKK,JKM'	'JKK,JHT,JKM'
588	'JKK,JKM'	'JKK,JKM'
133	'JHT,JKM'	'JKK,JHT,JKM'
762	'JKK,JHT,JKMLP'	'JKK,JHT,JKMLP'
652	'JHT,JP'	'JHT,JKMLP'
352	'JKK,JHT'	'JHT,JKM'
313	'JHT,JKM'	'JKK,JHT,JKM'
275	'JKK,JHT,JKM'	'JKK,JHT,JKM'
150	'JKK,JHT,JKM'	'JKK,JHT,JKM'
335	'JKK,JHT'	'JHT,JKM'
834	'JKK,JHT,JKMLP'	'JKK,JHT,JKMLP'
234	'JHT,JKM'	'JKK,JHT,JKM'
70	'JHT,JKM'	'JHT,JKM'
495	'JKK,JHT,JKM'	'JKK,JHT,JKM'
123	'JKK,JHT,JKM'	'JKK,JHT,JKM'
403	'JHT,JKM'	'JKK,JHT'
92	'JKK,JKM'	'JKK,JHT,JKM'
721	'JKK,JHT,JKM'	'JKK,JHT,JKM'
343	'JHT,JKM'	'JKK,JHT,JKM'
567	'JKK,JHT,JKMLP'	'JKK,JHT,JKM'
655	'JHT,JKM'	'JKK,JHT,JKM'
653	'JKK,JKM'	'JKK,JHT,JKM'
701	'JKK,JHT,JKM'	'JKK,JHT,JKM'
302	'JHT,JKM'	'JKK,JHT,JKM'
86	'JKK,JHT,JKM'	'JKK,JKM'
634	'JKK,JKM'	'JKK,JHT,JKM'
267	'JKK,JHT,JKM'	'JKK,JHT,JKM'
Jumlah data yang cocok		27

Gambar 11 Hasil Perbandingan Perhitungan Program MATLAB KNN dan Hasil Asli

Dari hasil gambar diatas kemudian dilakukan perhitungan akurasi dengan menggunakan Confusion Matrix sebagai berikut.

$$Akurasi = \frac{27}{50} \times 100\% = 54\%$$

Beberapa data yang tidak sesuai dengan dengan data asli ada 23 data yaitu sebagai berikut.

Data ke	Fakta Produk yang Dibeli	Hasil Program dengan k=7
265	'JHT, JKM'	'JKK, JHT, JKM'
138	'JHT, JKM'	'JKK, JHT'
133	'JHT, JKM'	'JKK, JHT, JKM'
313	'JHT, JKM'	'JKK, JHT, JKM'
234	'JHT, JKM'	'JKK, JHT, JKM'
403	'JHT, JKM'	'JKK, JHT'
343	'JHT, JKM'	'JKK, JHT, JKM'
655	'JHT, JKM'	'JKK, JHT, JKM'
302	'JHT, JKM'	'JKK, JHT, JKM'
652	'JHT, JP'	'JHT, JKM,JP'
563	'JKK, JHT, JKM,JP'	'JKK, JHT, JKM'
567	'JKK, JHT, JKM,JP'	'JKK, JHT, JKM'
892	'JKK, JHT, JKM'	'JKK, JHT, JKM,JP'
86	'JKK, JHT, JKM'	'JKK, JKM'
293	'JKK, JHT'	'JKK, JHT, JKM'
352	'JKK, JHT'	'JHT, JKM'
335	'JKK, JHT'	'JHT, JKM'
482	'JKK, JKM'	'JHT, JKM'
466	'JKK, JKM'	'JKK, JHT, JKM'
507	'JKK, JKM'	'JKK, JHT, JKM'
92	'JKK, JKM'	'JKK, JHT, JKM'
653	'JKK, JKM'	'JKK, JHT, JKM'
634	'JKK, JKM'	'JKK, JHT, JKM'

Gambar 12 Hasil Ketidakcocokan

Dari gambar diatas akan dilakukan analisa mengapa bisa terjadi ketidakcocokan dari hasil perhitungan yang ada dalam program MATLAB KNN. Dapat dilihat bahwa data yang mengalami hasil ketidakcocokan tersebut memiliki pola, hal ini ditunjukkan lewat tabel dengan perbedaan warna yang ditunjukkan dalam tabel. Data dengan label 845 dalam data asli teridentifikasi sebagai label 1043, untuk mengetahui mengapa hal ini dapat terjadi maka akan dilakukan analisis guna untuk mengoreksi hasil yang kurang valid tersebut.

Data dengan label 845 terdiri dari data ke 892 dan data 123, data tersebut memiliki variable seperti pada tabel 4.9 berikut ini.

Tabel 1 Tabel Data ke 597 dan 86

Data ke	Sex	Lahir	Upah
892	76	1986	2123500
123	76	1988	2018000

Dengan label yang sama akan kita analisis alasan mengapa data tersebut diklasifikasikan kedalam program 'JKK, JHT, JKM,JP'. Berikut merupakan tabel yang mempresentasikan data dengan label yang sama dengan data yang mengalami ketidakcocokan hasil.

Data ke	Sex	Lahir	Upah	Program
871	76	1988	2123500	'JKK, JHT, JKM,JP'
896	76	1986	2123500	'JKK,JKM'
898	76	1986	2123500	'JKK, JHT, JKM,JP'
902	76	1988	2123500	'JKK, JHT, JKM,JP'
913	76	1988	2123500	'JKK, JHT, JKM'
917	76	1986	2123500	'JHT, JKM,JP'
921	76	1988	2123500	'JKK, JHT, JKM,JP'
926	76	1988	2123500	'JKK, JHT, JKM,JP'
127	76	1986	2018000	'JKK, JHT, JKM'
135	76	1986	2018000	'JKK, JHT, JKM'

Gambar 13 Tabel Data dengan Label yang Sama

Dari tabel diatas dapat dilihat bahwa dengan label yang sama dan menerapkan nilai k yang telah ditentukan yaitu 5 maka ada 5 data yang terpilih paling dekat. Dari ke 5 data terpilih dalam data tersebut data dengan program 'JKK, JHT, JKM,JP' muncul 4kali, yaitu pada data ke 871, 898, 902 dan 921. Hal inilah yang menyebabkan program memprediksi data uji tersebut ke dalam kelas program 'JKK, JHT, JKM,JP'.

Selanjutnya kita juga dapat melihat bahwa dengan label yang sama data dengan kelas program 'JKK, JHT, JKM,JP' memiliki jumlah yang lebih banyak dibandingkan dengan data yang berkelas program 845. Data dengan kelas program 'JKK, JHT, JKM,JP' juga merupakan data yang muncul lebih awal dari data yang memiliki kelas program 845 dalam urutan data pelanggan sehingga program akan secara otomatis mencari data paling dekat dengan data uji, yaitu kelas program 'JKK, JHT, JKM,JP'. Sehingga data yang harusnya berkelas program JKK, JHT, JKM diprediksi oleh program dengan kelas program (JKK, JHT, JKM, JP). Mengingat bahwa konsumen dapat memiliki potensi yang beragam dalam menentukan produk asuransi, hal ini dapat memicu berubahnya hasil prediksi dalam menentukan produk asuransi.

4. KESIMPULAN

Dari penelitian yang telah dilakukan dapat ditarik kesimpulan sebagai berikut:

1. Berdasarkan perhitungan yang dilakukan dari program MATLAB KNN menunjukkan bahwa prediksi terhadap produk asuransi yang akan dipromosikan terhadap pelanggan dapat dilakukan dengan menggunakan algoritma KNN dan perhitungan jarak Manhattan Distance terhadap data pelanggan yang sudah ada sebelumnya. Penentuan produk yang akan dipromosikan dapat diklasifikasikan berdasarkan label-label yang telah ditentukan seperti yang sudah dibahas dalam bab 4.
2. Hasil akurasi dari program MATLAB KNN yaitu 54% menunjukkan bahwa hasil prediksi minat konsumen terhadap produk asuransi teruji. Seperti halnya yang dijelaskan dalam bab 4 bahwa hasil prediksi yang tidak cocok bukanlah sebuah hasil yang salah melainkan seorang konsumen dapat memiliki berbagai potensi dalam mengambil keputusan untuk membeli produk asuransi.

5. SARAN

Saran yang dapat diberikan kepada pihak BPJS Kota Semarang Cabang II adalah sebagai berikut:

1. Diharapkan BPJS untuk selalu melakukan analisa terhadap minat konsumen terhadap produk asuransi. Karena minat konsumen dapat berubah ubah, seperti yang dijelaskan dalam bab 4 bahwa konsumen dapat merubah pilihan mereka terhadap produk asuransi disaat muncul produk asuransi dari perusahaan. Mereka akan mempertimbangkan untuk membeli produk tersebut sesuai dengan kebutuhan mereka.
2. BPJS perlu melakukan inovasi baru guna untuk meningkatkan penjualan yang bertujuan untuk meningkatkan penjualan produk asuransi. Karena dengan adanya produk asuransi baru kemungkinan konsumen untuk membeli produk asuransi juga meningkat. Promosi penawaran produk yang tidak tepat terhadap pelanggan dapat menyebabkan pelanggan yang akan melakukan pembelian produk kemudian membatalkannya. Sehingga perusahaan perlu melakukan penawaran produk terhadap pelanggan dengan sangat hati hatiberbelanja dan pihak alfamart juga tidak mengalami kerugian karena pembeli yang pergi karena ketidak nyamanan proses belanja yang dilakukannya. Selain itu saran yang dapat diberikan bagi penelitian ini adalah, pengembangan program agar lebih *user friendly*, sehingga user tidak perlu lagi mengotak-atik bagian dalam program akan tetapi hanya perlu berhadapan dengan *user interface-nya*. Walaupun program ini sudah dapat dikatakan *user friendly* akan tetapi alangkah baiknya bila ada yang lebih baik lagi.

DAFTAR PUSTAKA

- [1] D. B. George Aloysius, "An approach to products placement in supermarkets using," *Journal of King Saud University – Computer and Information Sciences*, vol. 25, no. 1, pp. 77-87, 2013.
- [2] "Albion Research Ltd.," [Online]. Available: http://albionresearch.com/data_mining/market_basket.php. [Diakses 18 November 2015].
- [3] Y. R. B. A. O. D. Yessica Nataliani, "Market Basket Analysis dengan Metode Fuzzy C-Covering untuk Menenntukan Pola Pembelian pada Toko Buku".
- [4] Erwin, "Analisis Market Basket Dengan Algoritma Apriori dan FP-Growth".
- [5] I. S. d. A. C. Seruni, "DESIGN OF PRODUCT PLACEMENT LAYOUT IN RETAIL SHOP," *MAKARA, TEKNOLOGI*, vol. 9, pp. 43-47, 2005.
- [6] P. P. J. d. M. Mourya, "A Study on Market Basket Analysis Using a Data Mining," vol. 3, no. 6, 2013.
- [7] K. E. R. N. P. J. E. L. Mark J. Arnold, "Customer delight in a retail context: investigating delightful and terrible shopping experiences," *Journal of Business Research*, vol. 58, no. 8, pp. 1132-1145, 2005.
- [8] "Wikipedia," [Online]. Available: <http://id.wikipedia.org/wiki/Supermarket>. [Diakses 9 November 2015].
- [9] Alfamart, "Alfamart," Alfamart, 2013. [Online]. Available: <http://corporate.alfamartku.com/milestone>. [Diakses 30 june 2016].
- [10] B. Alma, *Manajemen Pemasaran dan Pemasaran Jasa*, Bandung: CV Alvabeta, 2009.
- [11] P. K. d. G. Armstrong, *Prinsip-Prinsip Pemasaran*. Edisi 13, Jilid 1, Jakarta: Erlangga, 201.
- [12] M. team, "Management Study Guide," [Online]. Available: <http://www.managementstudyguide.com/what-are-planograms.htm>. [Diakses 26 November 2015].
- [13] "Megeputer," [Online]. Available: http://www.megaputer.com/company/cases/cambridge_mba.php3. [Diakses 18 november 2015].
- [14] u. Fayyad, *Advance in Knowledge iscovery and Data Mining*, MIT Press, 1996.

- [15] W. a. Frank, *Data Mining: Practical Machine Learning Tools and Techniques Second Edition*, San Francisco: Morgan Kauffman, 2005.
 - [16] E. T. L. Kusriani, *Algoritma Data Mining*, Yogyakarta: Andi Offset, 2009.
 - [17] B. A. Mohammed Al-Maolegi, "AN IMPROVED APRIORI ALGORITHM FOR," *International Journal on Natural Language Computing (IJNLC)*, vol. 3, 2014.
 - [18] Y. Wicaksono, *Membangun Bisnis Online dengan Mambo*, Jakarta: PT. Elex Media Komputindo, 2008.
 - [19] Anhar, *Panduan Menguasai PH dan Mysql*, Jakarta: Media Kita, 2010.
 - [20] M. Arief, *Pemrograman Web Dinamis Menggunakan Php dan Mysql*, Yogyakarta: ANDI, 2011.
 - [21] A. Firmansyah, "IlmuKomputer.Com," [Online]. Available: <http://ilmukomputer.com/>. [Diakses 28 May 2016].
 - [22] "pengertian autocad," [Online]. Available: <http://www.ilmucad.com/2015/06/pengertian-autocad.html>. [Diakses 28 may 2016].
-