

**KLASIFIKASI STATUS KESEJAHTERANAAN RUMAH TANGGA
MENGUNAKAN ALGORITMA K – NEAREST NEIGHBOR DAN SELEKSI
FITURE BERBASIS CHI SQUARED**

*Indu Indah Purnomo, S.Kom, M.Kom
(indumbc@gmail.com)*

ABSTRAK

Kesejahteraan merupakan aspek penting dari kualitas manusia secara keseluruhan. Berbagai penelitian yang telah dilakukan mengenai kesejahteraan banyak sekali faktor yang mempengaruhi kesejahteraan rumah tangga. Oleh karena itu, dilakukan penelitian mengenai klasifikasi kesejahteraan rumah tangga di kecamatan Aluh aluh di Kota Banjarmasin. Algoritma k-Nearest Neighbor (k-NN) adalah suatu metode yang menggunakan algoritma supervised, dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada k-NN. Penelitian ini dilakukan untuk mengkaji hasil akurasi yang lebih tinggi dan kemudian mengaplikasikan Algoritma k-NN dalam klasifikasi data. Data yang digunakan hasil PPLS 2011 dimana data tersebut adalah data yang menunjukkan 2 katagori yaitu katogori miskin dan paling miskin. Pada seleksi Chi Squared berpengaruh terhadap kesejahteraan rumah tangga di Kecamatan Aluh aluh di Kota Banjarmasin.

Nilai K yang digunakan adalah hasil dari klasifikasi perhitungan menggunakan rapid miner. Hasil nilai K di hitung Mulai dari K=5 sampai k=15, Dengan 16 variable yang di hitung, dimana k =6 lebih tinggi akurasi 76.85% hasil akurasinya, dan katogori yang miskin 76.47% > 83.45% yang paling miskin 55.00 % > 70.39%

Kata Kunci : Algoritma klasifikasi k-Nearest Neighbor (k-NN), chi square, rapid miner.

PENDAHULUAN

Kemiskinan sering kali dipahami sebagai gejala rendahnya tingkat kesejahteraan semata, padahal kemiskinan merupakan gejala yang bersifat kompleks dan multidimensi, dimana berkaitan dengan aspek sosial, ekonomi, budaya, dan aspek lainnya. Berbagai upaya dan kebijakan pembangunan telah dilakukan pemerintah untuk menanggulangi kemiskinan. Seperti inpres desa tertinggal, pemberian BLT (Bantuan Langsung Tunai), raskin, kompensasi

BBM, dan berbagai program penanggulangan kemiskinan lainnya. Namun, dari berbagai program yang telah dilaksanakan oleh pemerintah tersebut, masih terdapat kekurangan-kekurangan dalam pelaksanaannya penelitian ini penulis ingin menerapkan algoritma yang berbeda, untuk mengetahui sampai sejauh mana tingkat akurasi dari kesejahteraan Rumah Tangga.

RUMUSAN MASALAH

- 1 Berdasarkan kondisi seperti yang telah diuraikan diatas, maka dapat dirumuskan permasalahan dalam penelitian ini yaitu *“Menganalisa kesejahteraan rumah tangga di Kecamatan Aluh-Aluh yang akurat menggunakan metode K-Nearest Neighbor future Chi squared*
- 2 Nilai akurasi kesejahteraan rumah tangga 76.85 % pada data set PPLS 2011

TUJUAN PENELITIAN

- 1 Berdasarkan latar belakang dan rumusan masalah diatas, maka penelitian ini bertujuan untuk menerapkan metode *K-Nearest Neighbor* melakukan klasifikasi status kesejahteraan RTS (Rumah Tangga Sasaran) Keluarga Binaan Sosial sehingga didapatkan tingkat akurasi pada data hasil PPLS 2011.
- 2 untuk meningkatkan akurasi pada kesejahteraan rumah tangga lebih besar dari 76.85% dengan pendekatan metode K – Nearest Neighbor berbasis Chi Square

TARGET LUARAN DAN MANFAAT PENELITIAN

Target luaran dalam pembuatan laporan penelitian ini adalah untuk hasil yang diharapkan dapat digunakan sebagai masukan bagi masarakat, khususnya membantu pemerintah dalam memberikan bantuan untuk keluarga yang tidak mampu

Manfaat bagi ilmu pengetahuan dan teknologi, hasil ini diharapkan dapat memberikan sumbangan berupa model metode k- nearest neighbor berbasis chi square pada kesejahteraan rumah tangga.

METODE PENELITIAN

1. Pengumpulan Data

Untuk mendukung pembuatan program ini dilakukan pengumpulan data dengan mengumpulkan bahan-bahan dari beberapa sumber, seperti media internet, buku-buku, terjun langsung kelapangan melihat dan mendata Kesejahteraan rumah tangga

2. Analisis Kebutuhan dan Perancangan Sistem

Merancang program dengan menggunakan Algoritma K- Nearest neighbor yang mudah dioperasikan. Metode perancangan sistem disusun berdasarkan hasil olah data yang diperoleh. Perancangan sistem dilakukan dengan menerapkan metode - metode yang didapat dari hasil observasi dan studi pustaka. Dalam metode ini hasil dari program yang di buat akan di tampilkan dalam bentuk tabel. Hasil laporannya dari input dan output yang akan di hasilkan.

3. Pembuatan Program

Mengimplementasikan hasil rancangan program kedalam bahasa pemograman. Klasifikasi perangkat lunak yang digunakan adalah Sistem operasi. Windows 7 dan tampilan algoritma K-Nearest neighbor (KNN).

4. Pengujian Program

Menguji dan menganalisis program, berdasarkan input dari pengguna serta keakuratan output yang dihasilkan. Percobaan untuk menjalankan sytem yang telah dibuat apakah sesuai dengan yang diinginkan. Pengujian sistem dengan menggunakan Algoritma K-Nearest neighbor (KNN).

ANALISA HASIL DAN PEMBAHASAN

1. Hasil Penelitian

Penelitian ini merupakan penelitian eksperimen, yakni penelitian yang berfokus pada penyelidikan beberapa variabel dan cara variabel-variabel tersebut dapat terpengaruh oleh suatu kondisi tertentu. Biasanya penelitian eksperimen digunakan untuk memverifikasi hipotesis yang sudah dirumuskan sebelumnya (Berndtsson, 2008). Dalam penelitian ini dilakukan penerapan Algoritma K – Nearest neighbor untuk analisa status kesejahteraan rumah tangga

2 Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah berasal dari hasil pendataan PPLS 2011 Provinsi Kalimantan Selatan Kecamatan Aluh – Aluh. Total data anggota rumah tangga di kecamatan Aluh – Aluh sebesar 18.434 data. Namun dalam penelitian hanya digunakan sebanyak 9.909 data dengan objek yang diteliti adalah status kesejahteraannya yang termasuk dalam kategori (1) miskin dan (2) paling miskin, yang usia kepala keluarganya antara 31 – 98 tahun, pengambilan sample penelitian ini dilakukan agar analisa bersifat independen antara satu dengan yang lain.

3 Metode Pengolahan Data Awal

Setelah pengumpulan data (Data Gathering) maka data tersebut kemudian diolah agar dapat diproses dalam data mining. Hal tersebut dilakukan karena dalam data mining, atribut yang kurang lengkap, tidak konsisten, dan tidak rapi tidak dapat ditangani secara sistematis oleh algoritma data mining. Oleh Karena itu itu perlu dilakukan proses validasi

untuk menemukan, dan mengkonversi data tersebut agar dapat digunakan dalam algoritma data mining. Data tersebut dapat ditangani dengan eliminasi, inspeksi, identifikasi, dan substitusi. Dalam dataset yang akan digunakan ini, validasi data yang digunakan adalah dengan mengeleminasi data yang tidak lengkap. Selanjutnya data yang tidak relevan dihilangkan dan data perolehan tersebut ditransformasi untuk mendapatkan atribut yang relevan dan sesuai dengan format input algoritma data mining.

4 Implementasi Sistem

Penelitian ini menghasilkan angka-angka yang dapat diukur dan nantinya akan dibandingkan untuk menentukan model yang cocok dan menghasilkan tingkat akurasi yang paling tinggi untuk klasifikasi status kesejahteraan RTS (Rumah Tangga Sasaran) Keluarga Binaan Sosial. Angka akurasi yang paling tinggi dihasilkan oleh suatu model akan menyimpulkan bahwa model tersebut cocok untuk melakukan klasifikasi dan menjadi salah satu faktor dalam penyimpulan model yang paling cocok digunakan.

5 Spesifikasi Perangkat Keras dan Perangkat Lunak

Program ini direkomendasikan untuk dijalankan dengan menggunakan perangkat keras (*hardware*) yang mempunyai spesifikasi berikut:

1. Prosesor Minimal Intel Pentium atau AMD.
2. Memory minimal 1 GB.
3. Harddisk minimal 320 GB.
4. VGA card 64 MB. (Optional)
5. Monitor dengan resolusi 1024 × 768 *pixel*.
6. *Keyboard* dan *Mouse*.


Adapun perangkat lunak (*software*) yang digunakan untuk menjalankan aplikasi ini adalah lingkungan sistem operasi minimal *MS-WindowXP* atau *Windows 7*.

6 Spesifikasi Program

Melakukan klasifikasi *K-Nearest Neighbor* tanpa seleksi atribut dapat diketahui hasil sehingga dapat dibandingkan di mana kategori 1 menunjukkan kategori kesejahteraan rumah tangga tersebut miskin dan kategori 2 menunjukkan kategori kesejahteraan rumah tangganya paling miskin.. Pada tahap validasi model klasifikasi dilakukan pengujian yang dilakukan terhadap data *testing* dengan metode tersebut dilakukan dengan teknik *folds cross validasi* dengan pengujian data mulai 2,3,4,5,6,7,8,9 dan 10 sehingga dapat di evaluasi hasilnya dengan mengukur seberapa keakuratan akurasi miskin dan paling miskin yang dihasilkan dari beberapa percobaan tersebut menggunakan metode *k-Nearest Neighbor*.

7 Pengujian Program

Dibawah ini adalah gambar – gambar hasil *accuracy* dari data kemiskinan pengujian validasi dengan pengujian data mulai dari 2 sampai 10 menggunakan metode *k- Nearest Neighbor* tanpa seleksi atribut.



Validasi	2	3	4	5	6	7	8	9	10
Accuracy	72.40	72.97	72.96	72.94	73.52	73.22	73.05	73.60	73.21
Precision	79.20	79.57	79.50	79.66	80.02	79.85	79.63	80.11	79.83
Recall	82.19	82.63	82.73	82.42	82.89	82.65	82.68	82.90	82.65
AUC	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500

Gambar 7.1 Hasil Tabel Pengujian KNN Tanpa seleksi Atribut

Dari hasil percobaan seperti gambar tabel 7.1 hasil pengujian sebanyak 10 kali percobaan dengan hasil pengukuran tersebut didapatkan akurasi yang lebih tinggi adalah pengujian yang ke 9 memiliki akurasi lebih tinggi dari yang lain sebesar 73.60 %.


8 Hasil Dari Pengujian Program

Hasil akurasi yang dihasilkan sebesar 73.60 % sudah cukup baik namun untuk meningkatkan nilai akurasi yang telah dihasilkan supaya lebih baik lagi maka peneliti mencoba melakukan seleksi atribut menggunakan *K – Nearest neighbor* berbasis *Chi Squared*.

Dari perolehan nilai akurasi data kemiskinan dari k 5 sampai k 15 dengan menggunakan metode *K- Neorest Neighbor* berbasis *future chi Squared*.

- A. Katogori kesejahteraan rumah tangga yang miskin 76.47 % > 83.45 %
- B. katogori kesejahteraan rumah tangga yang paling miskin 55.00% > 70.39%

Dari hasil gambar dibawah ini pengujian seleksi metode *k – Nearest neighbor* berbasis *chi Square* mencari akurasi dari k 5 sampai k 15 yang mana nilai seleksi lebih tinggi.



K	5	6	7	8	9	10	11	12	13	14	15
Accuracy	75.63	76.85	75.94	76.52	75.39	73.66	73.04	72.82	73.37	73.63	73.26
Precision	76.75	78.46	79.63	83.47	82.17	80.50	80.02	79.68	79.88	80.09	79.91
Recall	94.20	92.72	88.49	82.93	82.88	82.35	81.99	82.15	82.85	83.00	82.59
AUC	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500

Gambar 8.1 Hasil Pengujian Seleksi Atribut *K – Nearest Neighbor* Berbasis *Chi Squared*

Dari gambar 8.1 terlihat penggunaan seleksi atribut dengan nilai K6 atribut menghasilkan tingkat akurasi tertinggi sebesar 76.85 %. Hasil tersebut kemudian digunakan untuk membandingkan seberapa besar kenaikan tingkat akurasi dihasilkan menggunakan *K- Nearest neighbor* dengan melakukan seleksi atribut berbasis *Chi Square*

Hasil seleksi perbandingan validasi *K- Nearest neighbor* tanpa atribut dengan *K – Nearest Neighbor* menggunakan atribut berbasis *Chi Squared* sebagai berikut



Metode	<i>Naive Bayes</i>	<i>Naive</i>
	Tanpa Seleksi Atribut	<i>Bayes</i> Berbasis <i>Chi Squared</i>
<i>Accuracy</i>	73.60	76.85
<i>Precision</i>	80.11	78.46
<i>Recall</i>	82.90	92.72
<i>AUC</i>	0.500	0.500

Gambar 8.2 Perbandingan *Nearest neighbor* dengan *Naive Bayes* Berbasis *Chi Squared*

Pada gambar 8.2 terlihat bahwa penggunaan seleksi atribut dapat meningkatkan hasil akurasi *K- Nearest neighbor* dari 73.60 % menjadi 76.85 %. Meskipun kenaikan yang dihasilkan tidak terlalu besar yaitu hanya sebesar 3,25 % namun secara umum seleksi atribut menggunakan *K- Nearest neighbor* berbasis *Chi Squared* lebih baik dari pada *K- Nearest neighbor* tanpa menggunakan seleksi atribut.

PENUTUP

1.KESIMPULAN

1. Hasil klasifikasi menunjukkan bahwa terdapat perbedaan dan persamaan karakteristik kesejahteraan rumah tangga di kecamatan aluh-aluh di Kota Banjarmasin antara kelompok rumah tangga miskin dan paling miskin dalam aspek kependudukan, pendidikan, perumahan, ketenagakerjaan, sosial ekonomi rumah tangga, dan teknologi-formasi dan komunikasi.

2. Hasil klasifikasi akurasi menunjukkan bahwa pola hubungan kesejahteraan rumah tangga di Kecamatan aluh aluh di kota Banjarmasin dengan faktor-faktor yang mempengaruhinya.

A. Katogori kesejahteraan rumah tangga yang miskin 76.47 % > 83.45 %

B. katogori kesejahteraan rumah tangga yang paling miskin 55.00% > 70.39%

hasil tersebut disimpulkan bahwa *K-Nearest Neighbor* (*K-NN*) berbasis *Chi Squared* mengklasifikasi status kesejahteraan Keluarga Binaan Sosial lebih akurat dan juga dapat mengetahui hasil katogori miskin dan sangat miskin.

3. Hasil klasifikasi *K-Nearest Neighbor* (*K-NN*) berbasis *Chi square* menunjukkan bahwa pada 6 kali replikasi diperoleh nilai ketepatan klasifikasi terbesar, yaitu sebesar 76.85% akurasinya.

2. Saran

Berdasarkan hasil analisis dan pembahasan, didapatkan informasi bahwa kesejahteraan rumah tangga Di

Kecamatan aluh-aluh di Kota Banjarmasin pada tahun 2011 lebih banyak dipengaruhi oleh status pekerjaan utama kepala rumah tangga sehingga diharapkan pemerintah Kota Banjarmasin dapat mengoptimalkan berbagai program yang telah dicanangkan untuk memperluas kesempatan kerja, agar dapat menekan tingkat kemiskinan di kecamatan aluh-alur di Kota Banjarmasin

(Studi Kasus : Koperasi Simpan Pinjam) : Universitas Diponegoro Semarang.2012

DAFTAR PUSTAKA

- [1] Amelia Renggapratiwi. *Kemiskinan Dalam Perkembangan Kota Semarang Karakteristik Dan Respon Thesis. Semarang : Universitas Diponegoro.*2009.
- [2] B. P. Statistik. *Analisis Data Kemiskinan berdasarkan Data Pendataan Program Perlindungan Sosial (PPLS) 2011*, Jakarta: BPS.2012
- [3] Bagus Yuswantara. *Pengontrolan Kualitas Produksi Mebel Di PT. Majawana dengan Diagram Kontrol Multivariat Atribut Berdasarkan Jarak Chi-Square*, thesis.2013
- [4] Daimler Chrysler. *Cross Industry Standart For Data Mining*, CRISP-DM, 1996.
- [5] Erysurya Ningrum, Bambang Widjanarko Otok . *Klasifikasi Kesejahteraan Rumah Tangga di Kota Malang Dengan Pendekatan Bagging Regresi Logistik*, ITS Bandung.2012.
- [6] Emerensye S.Y. Pandie. *Sistem Informasi Pengambilan Keputusan Pengajuan Kredit Dengan Algoritma K- Nearest Neighbour* (Studi Kasus : Koperasi Simpan Pinjam) : Universitas Diponegoro Semarang.2012
- [7] Faturokhman, Molo dan Marcelinus *Kemiskinan dan Kependudukan di Pedesaan Jawa: Analisis data Susenas 1992*. Yogyakarta : Pusat Penelitian Kependudukan Universitas Gadjah Mada.1995.
- [8] Han, J.K. *Data mining: Concept and Technique*. San Fransisco: Morgan Kaufmann Publisher.2011
- [9] Juen Ling, I Putu Eka N Kencana, Tjokorda Bagus Oka. *Analisis Sentimen Menggunakan Classifier Dengan Seleksi Fitur Chi Square* : E-jurnal.2014.
- [10] Kusriani, Luthfi, E.T. *Algoritma Data mining*. Yogyakarta: Andi Offset.2009
- [11] Krisandi, Helmi, Bayu Prihandono. *Algoritma K-Nearest Neighbor Dalam Klasifikasi Data Hasil Produksi Kelapa Sawit Pada PT Minamas Kecamatan Parindu* : paper.2013
- [12] Larose D. *Discovering Knowledge in Data*. USA:John Wiley’s and Son ;2005.
- [13] Margianti. *Konsep Sistem Pengambilan Keputusan*.1994
- [14] Maning, C., Raghavan, P. & Schutze, H. *Introduction to Information Retrieval*. London: Cambridge University Press.2008
- [15] Mulyanto. *Kemiskinan dan Kebutuhan Pokok*. Jakarta : Rajawali.2012

- [16] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook edition 2*. Springer new York Dordrecht Heidelberg London.2010.
- [17] Rayner. *Proses Klasifikasi dengan K- Nearest Neighbor* : Paper.2010
- [18] Z. Zhu, Y.S. Ong, and M. Dash “Wrapper-filter feature selection algorithm using a memetic framework.,” *IEEE transaction on systems, man , and Cybernetics Society*, vol. 37, no. 1, pp.70-6. 2007.