PERBANDINGAN METODE KLASIFIKASI NAÏVE BAYES DAN K-NEAREST NEIGHBOR PADA ANALISIS DATA STATUS KERJA DI KABUPATEN DEMAK TAHUN 2012

SKRIPSI

Disusun Oleh :

RIYAN EKO PUTRI

24010210120042

JURUSAN STATISTIKA

FAKULTAS SAINS DAN MATEMATIKA

UNIVERSITAS DIPONEGORO

SEMARANG

2014

PERBANDINGAN METODE KLASIFIKASI NAÏVE BAYES DAN K-NEAREST NEIGHBOR PADA ANALISIS DATA STATUS KERJA DI KABUPATEN DEMAK TAHUN 2012

Oleh :

RIYAN EKO PUTRI

24010210120042

Sebagai Salah Satu Syarat untuk Memperoleh Gelar

Sarjana Sains pada Jurusan Statistika

JURUSAN STATISTIKA

FAKULTAS SAINS DAN MATEMATIKA

UNIVERSITAS DIPONEGORO

SEMARANG

2014

HALAMAN PENGESAHAN I

Judul : **Perbandingan Metode Klasifikasi Naïve Bayes dan K-Nearest Neighbor pada Analisis Data Status Kerja di Kabupaten Demak Tahun 2012**

Nama : Riyan Eko Putri

NIM : 24010210120042

Jurusan : Statistika

Telah diujikan pada sidang Tugas Akhir tanggal 20 Agustus 2014 dan dinyatakan lulus pada tanggal 10 September 2014.

Semarang, 10 September 2014

Mengetahui,

a.n. Ketua Jurusan Statistika

Sekretaris Jurusan Statistika

Fakultas Sains dan Matematika

Drs. Agus Rusgiyono, M.Si.

NIP.196408131990011001

Panitia Penguji Ujian Tugas Akhir

Ketua,

Dra. Dwi Ispriyanti, M.Si.

NIP 195709141986032001

HALAMAN PENGESAHAN II

Judul : **Perbandingan Metode Klasifikasi Naïve Bayes dan K-Nearest Neighbor pada Analisis Data Status Kerja di Kabupaten Demak Tahun 2012**

Nama : Riyan Eko Putri

NIM : 24010210120042

Jurusan : Statistika

Telah diujikan pada sidang Tugas Akhir tanggal 20 Agustus 2014.

Semarang, 10 September 2014

Pembimbing I	Pembimbing II
Dra. Suparti, M.Si	Rita Rahmawati, S.Si, M.Si
NIP. 196509131990032001	NIP. 198009102005012002

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Allah SWT karena berkat rahmat dan hidayah-Nya, penulis dapat menyelesaikan Tugas Akhir dengan judul **"Perbandingan Metode Klasifikasi Naïve Bayes dan K-Nearest Neighbor pada Analisis Data Status Kerja di Kabupaten Demak Tahun 2012".**

Penulis menyadari bahwa dalam penyusunan laporan ini tidak lepas dari bimbingan dan dukungan yang diberikan beberapa pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada:

1. Ibu Dra. Dwi Ispriyanti, M.Si. selaku Ketua Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro.
2. Ibu Dra. Suparti, M.Si. selaku dosen pembimbing I dan Ibu Rita Rahmawati, S.Si, M.Si. selaku dosen pembimbing II yang telah memberikan bimbingan, arahan, dan motivasi hingga terselesaikannya tugas akhir ini.
3. Seluruh Dosen Jurusan Statistika FSM Universitas Diponegoro yang telah memberikan ilmu yang sangat berguna.
4. Semua pihak yang telah membantu, yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari masih terdapat kekurangan dalam penulisan laporan ini. Oleh karena itu, penulis mengharapkan kritik dan saran dari pembaca. Semoga Tugas Akhir ini dapat bermanfaat bagi semua pihak.

Semarang, September 2014

Penulis

ABSTRAK

Jumlah penduduk yang besar di Indonesia erat kaitannya dengan status kerja penduduknya apakah menganggur atau bekerja dimana ketika tidak diimbangi dengan lapangan kerja yang tersedia dapat menyebabkan tingkat pengangguran yang tinggi. Digunakan dua metode untuk melakukan klasifikasi status kerja pada penduduk angkatan kerja di Kabupaten Demak tahun 2012 yaitu metode Naïve Bayes dan K-Nearest Neighbor. Naïve Bayes merupakan metode pengklasifikasian yang didasarkan pada penghitungan probabilitas sederhana, sedangkan K-Nearest Neighbor merupakan metode pengklasifikasian yang didasarkan pada perhitungan kedekatan jarak. Variabel yang digunakan dalam menentukan status kerja seseorang apakah menganggur atau bekerja yaitu jenis kelamin, status dalam rumah tangga, status perkawinan, pendidikan, dan umur. Pengklasifikasian status kerja dengan metode Naïve Bayes diperoleh keakurasian sebesar 94.09% dan dengan metode K-Nearest Neighbor diperoleh keakurasian sebesar 96.06%. Untuk mengevaluasi hasil klasifikasi digunakan perhitungan Press's Q dan APER. Berdasarkan hasil analisis, diperoleh nilai *Press's Q* yang menunjukkan bahwa kedua metode sudah baik dalam pengklasifikasian data status kerja di Kabupaten Demak. Berdasarkan perhitungan APER, pengklasifikasian data status kerja di Kabupaten Demak menggunakan metode K-Nearest Neighbor memiliki tingkat kesalahan yang lebih kecil dibandingkan dengan metode Naïve Bayes. Dari analisis tersebut dapat disimpulkan bahwa metode K-Nearest Neighbor bekerja lebih baik dibandingkan dengan Naïve Bayes untuk kasus data status kerja di Kabupaten Demak tahun 2012.

Kata kunci: Klasifikasi, Naïve Bayes, K-Nearest Neighbor (K-NN), Evaluasi klasifikasi

ABSTRACT

Large population in Indonesia is closely related to the working status of the population which is unemployed or employed. It can lead to the high unemployment when the avaliable jobs arent balance with the population. Used two methods to perform the classification of employment status on the number of residents in the labor force in Demak for 2012 which is Naïve Bayes and K-Nearest Neighbor. Naïve Bayes is a classification method based on a simple probability calculation, while the K-Nearest Neighbor is a classification method based on the calculation of proximity. Variables used in determining whether a person's employment status is idle or not are gender, status in the household, marital status, education, and age. Employment status of the data processing methods of Naïve Bayes with the accuracy obtained is equal to 94.09% and the K-Nearest Neighbor method obtained is equal to 96.06% accuracy. To evaluate the results of the classification used calculations Press's Q and APER. Based on the analysis, the Press's Q values obtained indicate that both methods are already well in the classification of employment status data in Demak. Based on the calculation of APER, the classification of data in the employment status of Demak using the K-Nearest Neighbor method has an error rate smaller than the Naïve Bayes method. From this analysis it can be concluded that the K-Nearest Neighbor method works better compared with the Naïve Bayes for employment status data in the case of Demak for 2012.

Keywords: Classification, Naïve Bayes, K-Nearest Neighbor (K-NN), Classification evaluation

DAFTAR ISI

Halaman

DAFTAR GAMBAR

Halaman

DAFTAR TABEL

Halaman

DAFTAR LAMPIRAN

Halaman

BAB I

LATAR BELAKANG

1.1 Latar Belakang

Indonesia merupakan negara yang luas dengan beribu pulau di dalamnya menyebabkan negara ini memiliki jumlah penduduk yang besar dengan karakteristik masyarakat yang bermacam-macam. Jumlah penduduk yang besar ini erat kaitannya dengan status kerja penduduknya apakah menganggur atau tidak menganggur (bekerja) dimana ketika tidak diimbangi dengan lapangan kerja yang tersedia dapat menyebabkan tingkat pengangguran yang tinggi. Pengangguran seringkali menjadi masalah yang krusial dalam perekonomian. Dengan adanya pengangguran, produktivitas dan pendapatan masyarakat akan berkurang sehingga dapat menyebabkan timbulnya kemiskinan dan masalah-masalah sosial lainnya. Dari data BPS diketahui bahwa pada tahun 2013 terdapat 7.39 juta penduduk yang tidak memiliki mata pencaharian dari total angkatan bekerja sebanyak 118.19 juta jiwa. Selain itu Tingkat Pengangguran Terbuka (TPT) di Indonesia pada Agustus 2013 mencapai 6.25 persen. Angka tersebut mengalami peningkatan dibanding TPT Februari 2013 yaitu sebesar 5.92 persen serta TPT Agustus 2012 sebesar 6.14 persen.

Pada kenyataannya dengan jumlah pengangguran yang relatif besar ini memberikan dampak negatif bagi berbagai sisi yang saling berkaitan. Menurut Pradana (2013), beberapa dampak sosial ekonomi dari tingkat pengangguran yang tinggi di antaranya, yang pertama adalah jumlah kemiskinan bertambah. Banyak keluarga tidak mampu memenuhi kebutuhan pokok seperti makanan, kesehatan,

pakaian maupun biaya pendidikan bagi anggota keluarganya. Hal ini mengakibatkan banyak anak putus sekolah dan memaksa mereka untuk menjadi anak jalanan ataupun bekerja di bawah umur demi memenuhi kebutuhan hidup. Dampak sosial ekonomi yang kedua adalah efek psikologis. Seseorang yang tidak mampu memenuhi kebutuhan menjadi tertekan dan stress, hal ini dapat mengubah pola pikir seseorang untuk melakukan tindakan kriminalitas ataupun premanisme sehingga tingkat keamanan di Indonesia menjadi berkurang. Dampak sosial ekonomi yang terakhir yaitu kesenjangan sosial masyarakat semakin tinggi. Akan banyak perumahan kumuh yang berada di sekitar gedung pencakar langit ataupun kota-kota metropolit karena persaingan hidup yang sangat ketat. Oleh karena itu dalam kasus ini penting dilakukan pengklasifikasian status kerja apakah menganggur atau tidak karena pengangguran merupakan salah satu faktor indikasi apakah suatu negara tersebut dikatakan sudah sejahtera atau belum.

Naïve Bayes dan K-Nearest Neighbor merupakan metode pengklasifikasi yang terkenal dengan tingkat keakuratan yang baik. Banyak penelitian telah dilakukan berkaitan dengan metode klasifikasi tersebut. Kebanyakan dari penelitian tersebut berbasiskan pada ilmu komputer atau informatika sehingga pada pembahasannya lebih ditekankan pada hasil pemrograman serta tema yang diambil berkaitan dengan hal-hal yang bersifat elektronik. Contoh dari penelitian sebelumnya yaitu artikel tentang klasifikasi email spam dengan menggunakan metode Naïve Bayes (Anugroho, 2010), klasifikasi SMS pada smartphone android dengan menggunakan metode Naïve Bayes (Ebranda dan Triana, 2013), serta klasifikasi dokumen berbahasa Indonesia dengan menggunakan metode K-Nearest Neighbor (Ridok, 2010). Metode Naïve Bayes dan K-Nearest Neighbor juga

merupakan metode pengklasifikasian yang cocok digunakan pada data dengan kelas Y bertipe kategorik dimana untuk data status kerja kelas yang digunakan yaitu pengangguran dan bukan pengangguran. Selain itu berbeda dengan metode pengklasifikasian dengan regresi logistik ordinal maupun nominal, pada metode Naïve Bayes dan K- Nearest Neighbor pengklasifikasian tidak diperlukan adanya permodelan maupun uji statistik seperti uji signifikansi.

Naïve Bayes merupakan metode pengklasifikasian peluang sederhana dengan asumsi antar variabel penjelas saling bebas (independen). Naïve Bayes dapat digunakan untuk berbagai macam keperluan antara lain untuk klasifikasi dokumen, deteksi spam atau filtering spam, dan masalah klasifikasi lainnya. K-Nearest Neighbor atau dapat disingkat dengan K-NN adalah salah satu metode non parametrik yang digunakan dalam pengklasifikasian. Metode K-NN pertama kali diterapkan pada awal 1950. K-NN merupakan jajaran metode sederhana yang sering disebut dengan *Lazy Learning*. Pada penulisan tugas akhir kali ini akan diaplikasikan kedua metode tersebut pada bidang statistika dengan permasalahan yang diangkat adalah kependudukan serta membandingkan keoptimalan dua metode tersebut dalam mengklasifikasi data status kerja di Kabupaten Demak pada tahun 2012.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka dapat dirumuskan permasalahan yaitu seberapa tepat hasil perbandingan pengklasifikasian pada metode Naïve Bayes dan K-Nearest Neighbor sesuai dengan data status kerja untuk Kabupaten Demak pada tahun 2012.

1.3 Batasan Masalah

Permasalahan pada penelitian ini dibatasi untuk daerah Kabupaten Demak, sesuai dengan pendataan yang dilakukan oleh BPS pada tahun 2012. Pengolahan tersebut menggunakan dua metode, yaitu metode Naïve Bayes dan K-Nearest Neighbor.

1.4 Tujuan

Melakukan klasifikasi, mengetahui akurasi klasifikasi, serta membandingkan hasil klasifikasi pada data status kerja di Kabupaten Demak tahun 2012 dengan menggunakan metode Naïve Bayes dan K-Nearest Neighbor (K-NN).