**LAPORAN TUGAS BESAR DATA MINING TENTANG PENERAPAN METODE KNN UNTUK PENETUAN PENERIMA BEASISWA PT. POS INDONESIA DI POLITEKNIK POS INDONESIA**

Diajukan Untuk Memenuhi Tugas Matakuliah Data Mining

Pada Program Studi D4 Teknik Informatika

****

Oleh:

Fikri Aldi Nugraha 1.16.4.038

Lalita Chandiany Adiputri 1.16.4.043

**PROGRAM DIPLOMA IV TEKNIK INFORMATIKA**

**POLITEKNIK POS INDONESIA**

**BANDUNG**

**2019**

# **KATA PENGANTAR**

Puji Syukur kehadirat Allah SWT yang telah memberikan rahmat dan hidayah-Nya, sehingga penulis mampu menyelesaikan laporan tugas besar yang berjudul **“Laporan Tugas Besar Data Mining Tentang Penerapan Metode Knn Untuk Penetuan Penerima Beasiswa Pt. Pos Indonesia Di Politeknik Pos Indonesia“** tepat pada waktunya. Laporan ini disusun untuk memenuhi kelulusan matakuliah Data Mining.

Penulis juga menyadari bahwa selama berlangsungnya penyusunan laporan ini tak lepas dari dukungan serta bantuan berbagai pihak. Kiranya tak ada ucapan yang lebih baik kecuali rasa terima kasih kepada :

1. M. Yusril Helmi Setyawan, S.Kom., M.Kom. selaku Ketua Program Studi DIV Teknik Informatika Politeknik Pos Indonesia.
2. Cahyo Prianto, S.Pd., M.T selaku Dosen pengampu yang telah memberikan bimbingan kepada penulis.
3. Teman-teman di Prodi Teknik Informatika, yang banyak memberikan semangat dan dukungan.
4. Serta semua pihak yang tidak dapat penulis sebutkan satu persatu.

Akhirnya, sebaris doa penulis panjatkan semoga hidayah Allah SWT, selalu di limpahkan kepada seluruh mahluk-Nya. Semoga laporan ini dapat bermanfaat bagi seluruh mahasiswa khususnya bagi penulis.

Bandung, 6 Juli 2019

Penulis

# **DAFTAR ISI**

[KATA PENGANTAR i](#_Toc13475689)

[DAFTAR ISI ii](#_Toc13475690)

[BAB I PENDAHULUAN I-1](#_Toc13475691)

[1.1 Latar Belakang I-1](#_Toc13475692)

[1.2 Identifikasi Masalah I-2](#_Toc13475693)

[1.3 Tujuan I-2](#_Toc13475694)

[BAB II LANDASAN TEORI II-3](#_Toc13475695)

[2.1 Beasiswa II-3](#_Toc13475696)

[2.2 PT.Pos Indonesia II-3](#_Toc13475697)

[2.3 Metode K-Nearest Neighbor (KNN) II-3](#_Toc13475698)

[2.4 Bahasa Pemrograman R II-4](#_Toc13475699)

[BAB III METODOLOGI PENELITIAN III-5](#_Toc13475700)

[3.1 Diagram Metodologi Penelitian III-5](#_Toc13475701)

[3.2 Tahapan-Tahapan Diagram Alur Metodologi Penelitian III-6](#_Toc13475702)

[3.2.1 Perencanaan Penelitian III-6](#_Toc13475703)

[3.2.2 Identifikasi Masalah III-7](#_Toc13475704)

[3.2.3 Menentukan Metode Penelitian III-7](#_Toc13475705)

[3.2.4 Pengumpulan Data III-7](#_Toc13475706)

[3.2.5 Implementasi dan Pengujian III-7](#_Toc13475707)

[3.2.6 Kesimpulan dan Saran III-8](#_Toc13475708)

[BAB IV DATA YANG DIPEROLEH IV-9](#_Toc13475709)

[4.1 Pengumpulan IV-9](#_Toc13475710)

[4.2 Jenis-Jenis Data IV-9](#_Toc13475711)

[4.3 Data Yang Diperoleh IV-10](#_Toc13475712)

[BAB V PENGOLAHAN DATA V-11](#_Toc13475713)

[5.1 Pengolahan Data V-11](#_Toc13475714)

[5.2 Hasil Pengolahan Data V-13](#_Toc13475715)

[BAB VI KESIMPULAN VI-16](#_Toc13475716)

[6.1 Kesimpulan VI-16](#_Toc13475717)

[DAFTAR PUSTAKA](#_Toc13475718)

# **BAB I PENDAHULUAN**

## **Latar Belakang**

Setiap lembaga pendidikan perguruan tinggi terdapat banyak jenis beasiswa yang ditawarkan kepada mahasiswa yang unggul atau berprestasi dan kurang mampu. Untuk mengantisipasi agar beasiswa PT. Pos tersalurkan kepada yang berhak, maka diperlukan suatu sistem untuk keputusan yang diambil. Metode klasifikasi algoritma k-nearest neighbor merupakan salah satu metode pengklasifikasian data yang memiliki konsistensi yang kuat, dengan cara mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama berdasarkan pencocokan bobot (Kusrini dan Lutfhi, 2009).[1] Algoritma ini lebih efektif dalam melakukan training data yang besar dan dapat menghasilkan data yang lebih akurat.

Konsep penelitian dengan algoritma K-nearest neighbor telah banyak dilakukan oleh peneliti-peneliti sebelumnya diantaranya yaitu, k-nearest neighbor merupakan teknik yang sangat sederhana, efisien dan efektif dalam bidang pengenalan pola, kategori teks, pengolahan objek dan lain-lain, karena kesedrhanaan pengolahannya dan mampu melakukan training data dalam jumlah yang besar (Bathia, 2010). Salah satu masalah dari algoritma ini adalah efek yang sama dari semua atribut yang terdapat pada data baru dan data lama dalam dataset pelatihan (Moradian dan Baraani, 2009). K-nearest neighbor mampu melakukan training pada dataset penyakit diabetes untuk melihat dampak negatif hilangnya nilai imputasi dan solusi untuk penyembuhan. Akurasi algoritma k-nearest Neighbor diatas ratarata (Christobel et al., 2013). K-nearest neighbor yang diterapkan pada data konsumen yang menggunakan jasa keuangan kredit kendaraan bermotor, pada penelitian ini k-nearest neighbor digunakan untuk penentuan resiko kredit kendaraan bermotor (Leidiyana, 2013).

Permasalahan yang sering timbul adalah banyaknya calon penerima beasiswa PT.Pos Indonesia yang tidak tepat sasaran, sehingga diperlukan sebuah sistem yang dapat mengklasifikasikan calon penerima beasiswa PT.Pos Indonesia berdasarkan data training yang diambil dari data mahasiswa penerima beasiswa PT.Pos Indonesia sebelumnya (dataset). Sehingga pihak perguruan tinggi bisa mengatasi sejak dini permasalahan tersebut. Penggunaaan teknik data mining dengan algoritma k-nearest neighbor diharapkan mampu memberikan informasi yang berguna tentang teknik klasifikasi calon penerima beasiswa PT.Pos Indonesia.

## **Identifikasi Masalah**

Dari latar belakang diatas masalah yang teridentifikasi Antara lain :

1. Bagaimana perhitungan metode KNN untuk penentuan penerima beasiswa PT.Pos di Politeknik Pos Indonesia.
2. Bagaimana mengimplementasikan metode KNN sebagai pendukung keputusan klasifikasi mahasiswa calon penerima beasiswa PT.Pos Indonesia.

## **Tujuan**

Tujuan dalam pembuatan sistem klasifikasi mahasiswa peerima beasiswa PT.Pos Indonesia yaitu :

1. Melakukan perhitungan KNN untuk penentuan penerima beasiswa PT.Pos Indonesia di Politeknik Pos Indonesia.
2. Mengetahui cara implementasi KNN pada sistem penentuan penerima beasiswa PT.Pos Indonesia di Politeknik Pos Indonesia.

# **BAB II LANDASAN TEORI**

## **Beasiswa**

Beasiswa adalah Bantuan yang membantu orang terutama yang masih sekolah atau kuliah agar mereka dapat cepat menyelesaikan tugasnya dalam rangka mencari ilmu pengetahuan hingga selesai. Bantuan ini bisasanya berbentuk dana atau menunjang biaya atau ongkos yang harus dikeluarkan oleh anak sekolah atau mahasiswa selama menempuh masa pendidikan di tempat belajar yang di inginkan [2].

## **PT.Pos Indonesia**

Pos Indonesia merupakan sebuah Badan Usaha Milik Negara ([BUMN](https://id.wikipedia.org/wiki/BUMN)) Indonesia yang bergerak di bidang layanan [pos](https://id.wikipedia.org/wiki/Pos). Saat ini, bentuk badan usaha Pos Indonesia merupakan Perseroan Terbatas dan sering disebut dengan PT. Pos Indonesia. Bentuk usaha Pos Indonesia ini berdasarkan Peraturan Pemerintah Republik Indonesia Nomor 5 Tahun 1995. Peraturan Pemerintah tersebut berisi tentang pengalihan bentuk awal Pos Indonesia yang berupa [perusahaan umum](https://id.wikipedia.org/wiki/Perusahaan_umum" \o "Perusahaan umum) (perum) menjadi sebuah perusahaan persero.[3]

Berdiri pada tahun 1746, saham Pos Indonesia sepenuhnya dimiliki oleh [Pemerintah Indonesia](https://id.wikipedia.org/wiki/Pemerintah_Indonesia" \o "Pemerintah Indonesia). Saat ini Pos Indonesia tidak hanya melayani jasa pos dan kurir, tetapi juga jasa keuangan, ritel, dan properti, yang didukung oleh titik jaringan sebanyak lebih dari 4.000 [kantor pos](https://id.wikipedia.org/wiki/Kantor_pos" \o "Kantor pos) dan 28.000 Agen Pos yang tersebar di seluruh wilayah Indonesia.[3]

## **Metode K-Nearest Neighbor (KNN)**

K-nearest neighbor merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. Nilai k yang terbaik untuk algoritma ini tergantung pada data, secara umumnya, nilai k yang tinggi akan mengurangi efek noise pada klasifikasi, tetapi membuat batasan antarasetiap klasifikasi menjadi lebih kabur. Ada banyak cara untuk mengukur jarak kedekatan

antara data baru dengan data lama (data training), diantaranya euclidean distance dan manhattan distance (city block distance), yang paling sering digunakan adalah euclidean distance [1].

Algoritma metode K-NN sangatlah sederhana, bekerja berdasarkan jarak terpendek dariquery instance ke training sample untuk menentukan K-NN-nya. Training sample diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan

fitur dari data.Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasitraining sample.Sebuah titik pada ruang ini ditandai kelas c jika kelas c merupakan klasifikasi yang paling banyak ditemui pada k buah tetangga terdekat dari titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan Euclidean Distance yang direpresentasikan pada persamaan 1 sebagai berikut: [4]

Dimana matriks D(a,b) adalah jarak skalar dari kedua vector a dan b dari matriks dengan ukuran *“d ”* dimensi [4].

## **Bahasa Pemrograman R**

R adalah bahasa pemrograman dan sistem perangkat lunak yang dirancang khusus untuk mengerjakan segala hal terkait komputasi statistik. Bahasa pemrograman ini pertama kali dikembangkan pada tahun 1993 oleh dua orang pakar statistik yaitu Ross Ihaka dan Robert Gentleman di Auckland University, New Zealand. Sampai saat ini, bahasa pemrograman R terus berkembang secara pesat seiring dengan semakin populernya terminologi “Big Data” dan meningkatnya kebutuhan perusahaan akan data scientist untuk mengolah dan menganalisis data di perusahaan tersebut sebagai dasar pengambilan kebijakan dan mengautomatisasi proses bisnis menjadi data-driven. Bahasa pemrograman seperti Python dan R telah menjadi pilihan utama bagi para peneliti maupun praktisi di bidang data science untuk mengolah dan menganalisis data baik itu untuk kepentingan penelitian maupun bisnis. Oleh karena itu, bagi seorang pemula di bidang data science, R merupakan bahasa pemrograman yang sangat recommended untuk dikuasai.[5]

# **BAB III METODOLOGI PENELITIAN**

## **Diagram Metodologi Penelitian**

Metodologi penelitian adalah sekumpulan peraturan, kegiatan, dan prosedur yang digunakan oleh pelaku suatu disiplin. Metodologi juga merupakan analisis teoritis mengenai suatu cara atau metode. Penelitian merupakan suatu penyelidikan yang sistematis untuk meningkatkan sejumlah pengetahuan, juga merupakan suatu usaha yang sistematis dan terorganisasi untuk menyelidiki masalah tertentu yang memerlukan jawaban[6]. Hakekat penelitian dapat dipahami dengan mempelajari berbagai aspek yang mendorong penelitian untuk melakukan penelitian. Setiap orang mempunyai motivasi yang berbeda, di antaranya dipengaruhi oleh tujuan dan profesi masing-masing. Motivasi dan tujuan penelitian secara umum pada dasarnya adalah sama, yaitu bahwa penelitian merupakan refleksi dari keinginan manusia yang selalu berusaha untuk mengetahui sesuatu. Keinginan untuk memperoleh dan mengembangkan pengetahuan merupakan kebutuhan dasar manusia yang umumnya menjadi motivasi untuk melakukan penelitian [7]. Berikut ini adalah alur dari metodologi penelitian yang dilakukan di dalam penelitian ini :



Gambar 3.1 Diagram Alur

1. **Tahapan-Tahapan Diagram Alur Metodologi Penelitian**

Tahapan dari Metode Penelitian yang akan dilakukan berdasarkan pada Diagram Alur Metodologi Penelitian di atas.

### **Perencanaan Penelitian**

Pada tahap ini adalah tahap awal yang akan dilakukan dalam melakukan sebuah penelitian karena pada tahap ini kita akan menentukan masalah apa yang akan kita jadikan penelitian dan menggunakan metode apakah penelitian tersebut.

### **Identifikasi Masalah**

Pada penelitian kali ini masalah yang akan diteliti oleh penulis adalah mengenai terhadap menentukan penerima beasiswa PT.Pos Indonesia di Politeknik Pos Indonesia.

### **Menentukan Metode Penelitian**

Pada Penelitian ini metode yang akan digunakan adalah metode k-nears neighbor (KNN)*.* Karena metode KNN merupakan metode pengklasifikasian data yang memiliki konsistensi yang kuat, dengan cara mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama berdasarkan pencocokan bobot dan lebih efektif dalam melakukan training data yang besar dan dapat menghasilkan data yang lebih akurat [8].

### **Pengumpulan Data**

Pengumpulan data merupakan metode yang difungsikan untuk memperoleh informasi-informasi atau data-data terhadap kasus yang menjadi permasalahan dalam laporan ini [9]. Hal yang paling perlu dibutuhkan oleh penulis adalah informasi-informasi mengenai metode yang digunakan dalam penelitian kasus ini, yaitu metode KNN. Data yang diperoleh dalam kasus ini yaitu Bagian Kemahasiswaan di Politeknik Pos Indonesia.

### **Implementasi dan Pengujian**

Implementasi dan pengujian merupakan metode terakhir yang digunakan setelah analisa data inputan dan perancangan rancang segmentasi dilakukan. Metode ini akan menjelaskan tentang penerapan jalannya rancang segmentasi yang telah dianalisa selanjutnya diimplementasikan dan dilakukan pengujian untuk mengetahui tingkat keberhasilan segmentasi yang telah ada. Implementasi pengembangan segmentasi ini akan dikembangkan pada spesifikasi hardware dan software berikut:

1. Perangkat keras

Processor : Core i3

Memori (RAM) : 2048 MB

1. Perangkat Lunak

Sistem operasi : Windows 10 Profesional 64-bit Operating System

Tools perancangan : R Studio

Sementara untuk tahapan pengujian yang akan dilakukan pada analisa data meliputi pengujian, berdasarkan jumlah K.

### **Kesimpulan dan Saran**

Tahapan kesimpulan dan saran merupakan akhir dari penelitian ini. Tahapan ini berisi tentang kesimpulan dari hasil-hasil penelitian dan pengujian yang telah dilakukan pada penelitian tugas ini.

# **BAB IV DATA YANG DIPEROLEH**

1. **Pengumpulan Data**

Menurut Gulo (2002:110) pengumpulan data merupakan aktivitas yang dilakukan guna mendapatkan informasi yang diperlukan dalam rangka mencapai tujuan dari suatu penelitian. Adapun tujuan penelitian adalah jawaban dari rumusan masalah ataupun hipotesis penelitian, untuk dapat menjawabnya diperlukan data atau informasi yang diperoleh melalui tahapan pengumpulan data. Informasi atau data mempunyai karakteristik yang berbeda beda sehingga membutuhkan metode yang berbeda-beda pula [10].

## **Jenis-Jenis Data**

Jenis-jenis data yang terdapat dalam metode K-Nearest Neighbor sebagai berikut :

1. Data Train

Data train sering disebut dengan data uji yang merupakan data yang sudah berkelas atau berlabel yangb digunakan untuk menghitung akurasi model klasifikasi yang dibentuk [11].

1. Data Set

Dataset adalah objek yang merepresentasikan data dan relasinya di memory. Strukturnya mirip dengan data di database. Dataset berisi koleksi dari datatable dan datarelation [12]. Selain itu data set juga sering disebut data latih yang merupakan data yang sudah ada sebelumnya berdasarkan fakta yang sudah terjadi (predefined class) [13].

## **Data Yang Diperoleh**

Data yang diperoleh dalam penelitian ini kita dapatkan dari bagian kemahasiswaan di Politeknik Pos Indonesia. Berikut data yang kami peroleh :

****

Gambar 4.1 Data Beasiswa PT.Pos Indonesia

Gambar 4.1 merupakan data beasiswa PT.Pos Indonesia. Data tersebut terdapat 15 mahasiswa yang mengikuti seleksi beasiswa PT.Pos Indonesia, dimana hanya 6 mahasiswa yang menerima beasiswa dan 9 mahasiswa tidak menerima beasiswa. Data tersebut didapatkan dari bagian kemahasiswaan Politeknik Pos Indonesia.

# **BAB V PENGOLAHAN DATA**

## **Pengolahan Data**

Pengolahan data atau data processing merupakan manipulasi data ke bentuk yang lebih informative atau berupa informasi. Informasi merupakan hasil dari kegiatan pengolahan suatu data dalam bentuk tertentu yang lebih berarti dari suatu kegiatan atau suatu peristiwa[14].Terdapat empat penggolongan alat pengolahan data yang bisa anda ketahui diantaranya peralatan manual atau alat sederhana untuk mengolah data dengan faktor terpenting adalah dalam penggunaan alat dengan tenaga tangan manusia[15]. Berikut pengolahan data untuk penerima beasiswa PT.Pos Indonesia menggunakan bahasa R :

library(class)

library(data.table)

library(ggplot2)

library(plyr)

options(scipen=999) # turn-off scientific notation like 1e+48

#baca dataset

data.beasiswa <- read.csv("Dataset.csv", header=T)

new.data.beasiswa <- data.beasiswa[,c(3,4,5,6)]

beasiswa <- data.table(new.data.beasiswa)

View(beasiswa)

#knn function

prediksi\_beasiswa <- cbind(beasiswa$IPK, beasiswa$Gaji\_Orang\_Tua, beasiswa$Jumlah\_Tanggungan)

head(prediksi\_beasiswa)

keterangan <- beasiswa$Hasil

head(keterangan)

databaru <- cbind(2.73, 3500000, 5)

databaru

prediksi\_beasiswa.knn <- knn(prediksi\_beasiswa, databaru, keterangan, k=3, prob = T)

prediksi\_beasiswa.knn

#Visualisasi

plot.beasiswa = data.frame(prediksi\_beasiswa, Keputusan = beasiswa$Hasil)

plot.beasiswa1 = data.frame(x = plot.beasiswa$X1,

y = plot.beasiswa$X2,

Keputusan = plot.beasiswa$Keputusan)

find\_hull = function(beasiswa) beasiswa[chull(beasiswa$X1, beasiswa$X2), ]

boundary = ddply(plot.beasiswa1, .variables = "Keputusan", .fun = find\_hull)

ggplot(plot.beasiswa, aes(X1, X2, color = Keputusan, fill = Keputusan)) +

geom\_point(size = 5) +

annotate("point", x = 2.73, y = 3500000, size = 5, colour="black") +

labs(subtitle="Klasifikasi mahasiswa penerima beasiswa",

y="Gaji Orang Tua",

x="IPK",

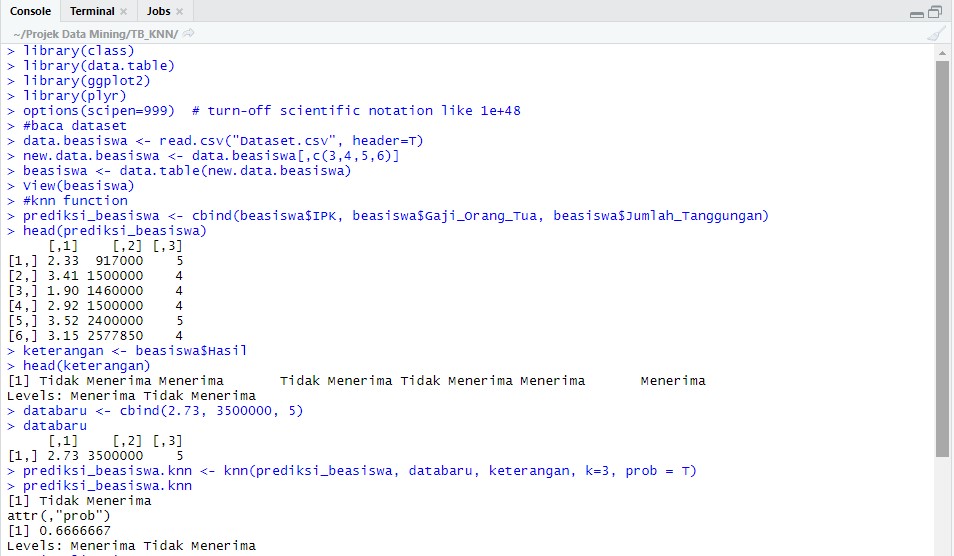
title="Penerima Beasiswa PT POS INDONESIA",

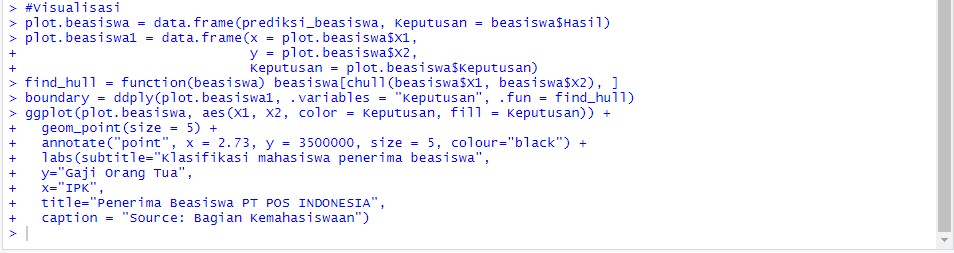
caption = "Source: Bagian Kemahasiswaan")

geom\_polygon(data = boundary, aes(x,y), alpha = 0.5)

## **Hasil Pengolahan Data**

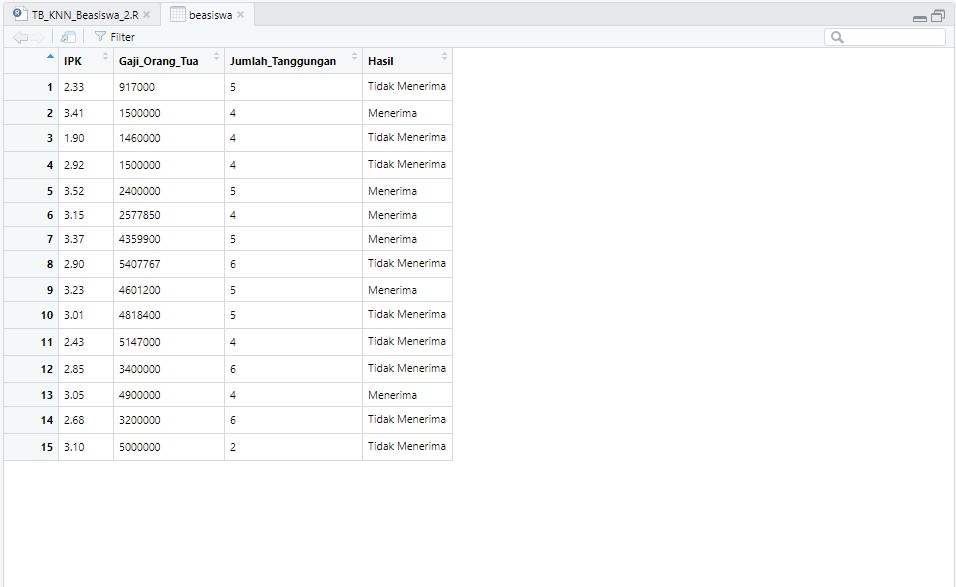
Hasil penerapan metode K- Nears Neighbor dengan menggunakan Bahasa R programming sebagai berikut:



****

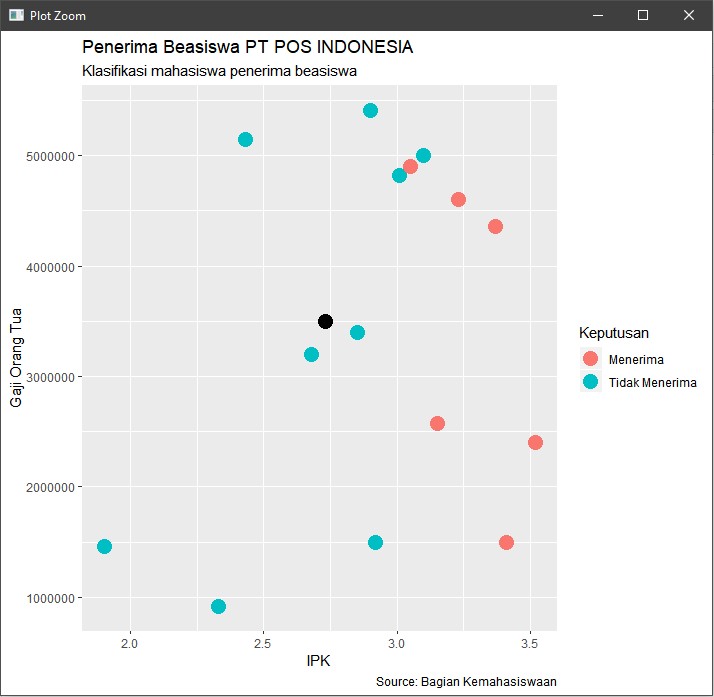
Gambar 5.1 Source Code Implementasi Metode KNN

Gambar 5.1 merupakan source code dari implementasi metode KNN dengan menggunakan Bahasa R . Pada gambar tersebut kita dapat mengetahui bahwa jumlah “k” yag kita gunakan yaitu k = 3 dengan data train yaitu ipk =2.73, penghasilan orangtua = 3.500.000 dan jumlah tanggungan = 5, maka diperoleh hasilnya “Tidak Menerima” dengan probabilitas 0,6666667.



Gambar 5.2 Tabel Data Mahasiswa

Gambar 5.2 merupakan gambar table data mahasiswayang di view melalui R Studio. Diamana data yang di tampilkan yaitu IPK, gaji orang tua, jumlah tanggungan dan hasil dengan jumlah data sebanyak 15 mahasiswa.

****

Gambar 5.3 Grafik Penerima Beasiswa PT. Pos Indonesia Menggunakan Metode KNN

Gambar 5.3 merupakan grafik penerima beasiswa PT. Pos Indonesia menggunakan metode KNN. Dalam grafik tersebut terdapat point dengan 2 warna dimana point yang bewarna merah merupakan mahasiswa yang menerima beasiswa dan point yang berwarna biru merupakan mahasiswa yang tidak menerima beasiswa. Selain itu didalam grafik tersebut adanya ipkpada garis horizontal (x) dan gaji orang tua pada garis vertical (y).

# **BAB VI KESIMPULAN**

## **Kesimpulan**

Berdasarkan hasil penerapan metode K- Nears Neighbor pada penentuan penerima beasiswa PT.Pos Indonesia di Politeknik Pos Indonesia dapat diambil kesimpulan sebagai berikut :

1. Metode K-Nears Neighbor merupakan metode merupakan metode klasifikasi yang dapat digunakan untuk memberikan keputusan alternatif dalam menentukan calon penerima beasiswa berdasarkan data-data yang telah diperoleh (data sampel).
2. Dengan menerapkan metode KNN untuk penentuan penerima beasiswa dengan K = 3 dan data trainnya meliputi ipk = 2.73, gaji orang tua = 3.500.000 , serta jumlah tanggung = 5 hasilnya tidak diterima dengan probabilitas 0,6666667 perhitungan tersebut diolah menggunakan tools R studio.

# **DAFTAR PUSTAKA**

1. Sumarlin. "Implementasi Algoritma K-Nears Neighbor Sebagai Penduduk Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM." *Jurnal Sistem Informasi Bisnis 01.* 13 April  2015.
2. Hendri Risman, Didik Nugroho, dan Yustina Retno. "Penerapan Metode K-Nears Neighbor Pada Aplikasi Penentu Penerima Beasiswa Mahasiswa di STMIK Sinae Nusantara Surakarta". *Jurnal TIKomSiN*  (2018).
3. Wikipedia.2019. <https://id.wikipedia.org/wiki/Pos_Indonesia> (diakses 7 Juli 2019).
4. Jodi Irjaya Kartika, EdySantoso , Sutrisno. “ Penentuan Siswa Berprestasi Menggunakan Metode K-Nearest Neighbor dan Weighted Product (Studi Kasus : SMP Negeri 3 Mejayan)”. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* e-ISSN: 2548-964X Vol. 1, No. 5. Mei 2017.
5. Amanda Pratama Putra. 2018. <https://medium.com/@mandes95/belajar-data-science-langkah-awal-mengenal-r-dan-rstudio-198ec2246f78> (diakses 5 Juli 2019).
6. Caesar Rama Raharya, Nurul Hidayat dan Edy Santoso. “ Penentuan Penerimaan Beasiswa Menggunakan Metode Modified K-Nearest Neighbor”. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN: 2548-964X Vol. 2, No. 11*.November 2018.
7. Ni Luh Gede Pivin Suwirmayanti “ Penerapan Metode K-Nearest Neighbor Untuk Sistem Rekomendasi Pemilihan Mobil”. *Techno.COM, Vol. 16, No. 2*. Mei 2017
8. Henny Leidiyana “Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bemotor”*.* *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic (2013).*
9. Anwar Hidayat.2016.Pengertian dan Penjelasan Metodologi Penelitian Lengkap di <https://www.statistikian.com/2016/11/metodologi-penelitian.html> (diakses 5 Juli 2019).
10. Henny Leidiyana*.*” Penerapan Metode K-Nearest Neighbor Pada Penentuan Grade Dealer Sepeda Motor”. *Jurnalilmu Pengetahuan Dan Teknologi Komputer (2017).*
11. Adam Mukharil Bachtiar. 2017. Data Management (Data Mining Klasifikasi) di <https://www.slideshare.net/adfbipotter/data-management-data-mining-klasifikasi> (diakses 4 Juli 2019).
12. Muhammad Zakaria. 2018. Pengertian Data Beserta Fungsi Data dan Jenis-Jenis Data yang Wajib Anda Ketahui di <https://www.nesabamedia.com/pengertian-data/> (diakses 4 Juli 2019).
13. Prasetia Utama. 2016. Training dan Testing di <https://prasetiautamacv.wordpress.com/2016/07/10/training-dan-testing/> (diakses 3 Juli 2019).
14. Ulfah. 2016. Pengertian Pengolahan Data : Fungsi dan Contohnya di <http://sitimariyaulfah.blogspot.com/2016/12/pengertian-pengolahan-data-fungsi-dan.html> (diakses 3 Juli 2019).
15. Arif Ashkaf. 2014. Tentang Pengolahan Data di <https://arifashkaf.wordpress.com/2014/05/08/tentang-pengolahan-data/> (diakses 3 Juli 2019).