

**STUDI KOMPARASI METODE *NAIVE BAYES* DAN *K-NEAREST NEIGHBOR* (KNN) UNTUK MEMPREDIKSI PENDONOR DARAH POTENSIAL (STUDI KASUS: *BLOOD TRANSFUSION SERVICE CENTER*)**

Skripsi

Untuk memenuhi sebagai persyaratan

mencapai derajat Sarjana – S1

Program Studi Teknik Informatika



Disusun oleh

STATE ISLAMIC UNIVERSITY  
**SUNAN KALIJAGA**  
YOGYAKARTA

NUR FAIZAH

16650012

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA  
YOGYAKARTA**

2020



**PENGESAHAN TUGAS AKHIR**

Nomor : B-2057/Un.02/DST/PP.00.9/09/2020

Tugas Akhir dengan judul : **STUDI KOMPARASI METODE NAIVE BAYES DAN K-NEAREST NEIGHBOR (KNN) UNTUK MEMPREDIKSI PENDONOR DARAH POTENSIAL (STUDI KASUS : BLOOD TRANSFUSION SERVICE CENTER)**

yang dipersiapkan dan disusun oleh:

Nama : **NUR FAIZAH**  
Nomor Induk Mahasiswa : **16650012**  
Telah diujikan pada : **Senin, 31 Agustus 2020**  
Nilai ujian Tugas Akhir : **A-**

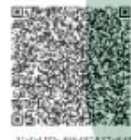
dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

**TIM UJIAN TUGAS AKHIR**



Ketua Sidang/Penguji I  
Maria Ulfah Siregar, S.Kom. MIT., Ph.D.  
SIGNED

Valid ID: 5f754f1db56ac



Penguji II  
Muhammad Didik Rohmad Wahyudi, S.T.,  
MT.  
SIGNED

Valid ID: 5b6d5f517c645



Penguji III  
Rahmat Hidayat, S.Kom., M.Cs.  
SIGNED

Valid ID: 5f7324cb58aac



Yogyakarta, 31 Agustus 2020  
UIN Sunan Kalijaga  
Dekan Fakultas Sains dan Teknologi  
Dr. Hj. Khurid Wanduti, M.Si.  
SIGNED

Valid ID: 5f767aa1b268b

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA



Wa

**SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR**

Hal : Persetujuan Skripsi

Lamp :

Kepada  
Yth. Dekan Fakultas Sains dan Teknologi  
UIN Sunan Kalijaga Yogyakarta  
di Yogyakarta

*Assalamu'alaikum wr. wb.*

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Nur Faizah  
NIM : 16650012  
Judul Skripsi : Studi Komparasi Metode *Naive Bayes* dan *K-Nearest Neighbor* (KNN) untuk Memprediksi Pendorong Darah Potensial. (Studi Kasus: *Blood Tranfusion Service Center*)

sudah dapat diajukan kembali kepada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Teknik Informatika

Dengan ini kami berharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqsyahkan. Atas perhatiannya kami ucapkan terima kasih.

*Wassalamu'alaikum wr. wb.*

Yogyakarta, 28 Agustus 2020  
Pembimbing

Maria Ulfah Siregar, S.Kom. MIT., Ph.D.  
NIP. 19780106 200212 2 001

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

## PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan dibawah ini:

Nama : Nur Faizah

NIM : 16650012

Program Studi : Teknik Informatika

Fakultas : Sains dan Teknologi

Menyatakan bahwa skripsi saya yang berjudul "**Studi Komparasi Metode Naive Bayes dan K-Nearest Neighbor untuk Memprediksi Pendorong Darah Potensial (Studi Kasus: Blood Tranfusion Service Data Center)**" merupakan hasil penelitian saya sendiri, tidak terdapat pada karya yang pernah di ajukan untuk memperoleh gelar kesarjana di suatu perguruan tinggi, dan bukan plagiasi karya orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 28 Agustus 2020

Yang menyatakan

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA



## KATA PENGANTAR

Alhamdulillah, segala puji bagi Allah SWT yang telah melimpahkan rahmat, hidayah, taufiq dan inayah-Nya kepada seluruh hamba-Nya. Shlawat serta salam senantiasa tercurah limpahkan kepada Nabi Muhammad SAW. Puji syukur penulis panjatkan kehadirat Allah SWT yang mana penyusunan skripsi ini akhirnya dapat diselesaikan. Penulis menyadari bahwa skripsi ini masih memiliki banyak kekurangan dan kesalahan. Oleh karena itu, penulis sangat mengharapkan saran dan kritik yang membangun agar dapat menghasilkan karya yang lebih baik lagi di kemudian hari. Proses penulisan skripsi ini tentu tidak terlepas dari bantuan dan dukungan dari berbagai pihak. Untuk itu penulis haturkan terimakasih yang sebesar-besarnya kepada:

1. Allah SWT. atas semua limpahan rahmat yang telah dianugerahkan dan kepada Nabi Muhammad SAW. Yang telah menunjukkan jalan kebenaran kepada umatnya.
2. Bapak Prof. Dr. Phil. Al-Makin, S.Ag., MA., selaku Rektor UIN Sunan Kalijaga Yogyakarta.
3. Ibu Dr. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
4. Ibu Maria Ulfah Siregar, S.Kom. MIT., Ph.D., selaku Ketua Program Studi Teknik Informatika UIN Sunan Kalijaga Yogyakarta.
5. Bapak Agung Fatwanto, selaku Dosen Pembimbing Akademik penulis dari semester awal hingga penulis menyelesaikan proses belajar di Prodi Teknik Informaika. Terimakasih telah memberikan arahan dan bimbingan

kepada penulis selama menuntut ilmu di Prodi Teknik Informatika hingga penulis dapat menyelesaikan skripsi ini.

6. Ibu Maria Ulfah Siregar, S.Kom. MIT., Ph.D., selaku Dosen Pembimbing Skripsi yang telah meluangkan waktu untuk membaca, mengoreksi dan membimbing penulis. Terimakasih banyak atas bimbingan serta motivasi dari Ibu. Banyak pelajaran dan pengetahuan yang penulis dapatkan selama bimbingan dengan Ibu.
7. Seluruh Dosen Prodi Teknik Informatika pada khususnya, dan semua Dosen Fakultas Sains dan Teknologi yang telah menginspirasi serta memberikan sumbangsih ilmu yang sangat bermanfaat dan berarti bagi penulis. Kepada segenap Staf Tata Usaha, karyawan fakultas Sains dan Teknologi, Staf Perpustakaan UIN Sunan Kalijaga dan seluruh karyawan dan pegawai UIN Sunan Kalijaga di berbagai lini, terimakasih atas bantuannya selama penulis menempuh studi di UIN Sunan Kalijaga sampai selesai di jenjang strata satu.
8. Terkhusus kepada kedua orang tuaku Bapak Slamet Riyadi dan Ibu Nasiyah yang penulis hormati dan *ta'dzimi*. Penulis sampaikan terimakasih atas doa, nasihat didikan, bantuan dorongan dan semangat baik lahir maupun batin serta kasih sayangnya yang tak pernah putus. Semoga Allah senantiasa melindungi, menganugerahkan rahmat dan ridhonya dan selalu senantiasa diberikan kesejahteraan dalam hidup. Semoga penulis bisa mewujudkan apa yang engkau berdua cita-citakan. Amin.

9. Abah KH. Fairuzi Afiq Dalhar, Ibu Nyai Hj. Siti Mukarromah, selaku pengasuh PP. Almunawwir Komplek Nurussalam dan seluruh *dzurriyah* simbah K.H. Dalhar Munawwir yang telah menjadi salah satu sumber inspirasi penulis untuk segera menyelesaikan skripsi ini.
10. Maryam, Iin, Sila, Mila dan Sila Terimakasih telah menjadi keluarga baruku di Jogja, *Rahayu* kalian.
11. Keluarga saudah (mba hana, mba olif, mba iluk, mba zizah, mba mela, salma, aah, nabila, anis, mba naffa, alfi, ifah, aini). Terimakasih selalu mengingatkan dan mendukung penulis untuk terus mengejar cita-cita tanpa meninggalkan kewajiban mengaji. Mereka adalah sahabat dunia akhirat penulis. Amin.
12. Segenap pengurus pondok pesantren Almunawwir yang mengajarkan arti mengabdikan untuk keberkahan. Terimakasih telah menginspirasi bagaimana menjadi orang yang hebat dalam berbagai aspek.
13. Teman seperjuangan, Dinda, Intan, Yayah dan Teman-teman Teknik Informatika 2016 yang telah banyak memberikan bantuan, dukungan dan motivasi yang membangun dalam menuntut ilmu.
14. Teman-teman KKN Bibis (Puji, Shohi, Ria, Arum, Khawari, Irsyad, Yoga) yang telah menjadi teman mengabdikan selama kurang lebih 60 hari.
15. Teman diskusi dan bermain saat-saat jenuh Gus Habib, Gus Alfian, Dyah Rahayu, Muflikhatul Hidayah terimakasih telah menemani perjalanan panjang sejak kecil sampai saat ini.

16. Seluruh TIM *Internship* PIRI CBS Yogyakarta, terimakasih telah menyediakan ruang untuk berbagi dan belajar bersama.

17. Serta semua pihak yang tidak dapat penulis sebutkan satu persatu.

Penulis berdo'a semoga bentuk dukungan yang telah diberikan oleh semua pihak terkait, tercatat sebagai amal baik dan kelak akan mendapatkan balasan yang berlipat dari Allah SWT.

Cilacap, 26 Agustus 2020

Penyusun

Nur Faizah

16650012



STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA



## HALAMAN PERSEMBAHAN

Skripsi ini saya persembahkan untuk  
almamater dan keluarga tercinta yang selalu  
memberikan dukungan kepada penulis untuk  
terus tumbuh menjadi pribadi yang lebih baik.



STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

## HALAMAN MOTTO

الأعمال صور قائمة و أرواحها جود سیر الإخلاص فيها

**Amal itu kerangka yang mati, dan  
ruhnya adalah keikhlasan yang ada padanya.**

~Ibnu Athailah

Setiap harapan harus ada tindakan.

~faizah



STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>PENGESAHAN TUGAS AKHIR.....</b>	<b>ii</b>
<b>SURAT PERSETUJUAN SKRIPSI / TUGAS AKHIR.....</b>	<b>iii</b>
<b>PERNYATAAN KEASLIAN SKRIPSI.....</b>	<b>iv</b>
<b>KATA PENGANTAR.....</b>	<b>v</b>
<b>HALAMAN PERSEMBAHAN .....</b>	<b>ix</b>
<b>HALAMAN MOTTO .....</b>	<b>x</b>
<b>DAFTAR ISI.....</b>	<b>xi</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiii</b>
<b>DAFTAR TABEL .....</b>	<b>xiv</b>
<b>INTISARI .....</b>	<b>xv</b>
<b>ABSTRACT .....</b>	<b>xvi</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah.....	3
1.5 Manfaat Penelitian.....	3
1.6 Keaslian Penelitian .....	4
1.7 Sistematika Penelitian .....	4
<b>BAB II LANDASAN TEORI DAN TINJAUAN PUSTAKA.....</b>	<b>6</b>
2.1 Tinjauan Pustaka .....	6
2.2 Landasan Teori .....	13
2.2.1 Donor Darah.....	13
2.2.2 Data Mining .....	15
2.2.3 <i>K-Nearest Neighbor</i> .....	22
2.2.4 <i>Naive Bayes</i> .....	25
2.2.5 <i>Z-Score Normalization</i> .....	26
2.2.6 <i>Min-Max Normalization</i> .....	27
2.2.7 Seleksi Fitur .....	30
2.2.8 <i>Confusion Matrix</i> .....	30

2.2.9	RFMTC .....	32
2.2.10	Python .....	33
<b>BAB III</b>	<b>METODE PENELITIAN .....</b>	<b>32</b>
3.1	Studi Pustaka .....	32
3.2	Pengumpulan Data .....	33
3.3	Preprocessing.....	33
3.4	Prediksi .....	33
3.5	Analisis dan Evaluasi .....	33
3.6	Kebutuhan Sistem.....	34
<b>BAB IV</b>	<b>HASIL DAN PEMBAHASAN .....</b>	<b>35</b>
4.1	Pengumpulan Data .....	35
4.2	Preprocessing.....	36
4.2.1	Seleksi Fitur .....	36
4.2.2	<i>Z-Score Normalization</i> .....	37
4.2.3	<i>MinMax Normalization</i> .....	41
4.3	Prediksi Metode KNN .....	43
4.4	Prediksi Metode <i>Naive Bayes</i> .....	46
4.5	Analisis dan Evaluasi .....	50
4.5.1	<i>Confusion Matrix</i> .....	50
<b>BAB V</b>	<b>PENUTUP.....</b>	<b>35</b>
5.1	Kesimpulan.....	35
5.2	Saran.....	35
<b>DAFTAR PUSTAKA</b>	<b>.....</b>	<b>56</b>
<b>LAMPIRAN</b>	<b>.....</b>	<b>59</b>
<b>CURRICULUM VITAE</b>	<b>.....</b>	<b>72</b>

## DAFTAR GAMBAR

Gambar 4. 3 Seleksi Fitur Korelasi Antar Variabel Independen .....	37
Gambar 4. 4 <i>Confusion Matrix Naive Bayes Minmax</i> RMT ( 20 : 80) .....	51
Gambar 4. 5 <i>Confusion Matrix Naive Bayes Z Score</i> RMT ( 30 : 70).....	51
Gambar 4. 6 <i>Confusion Matrix KNN Z Score</i> RMT ( 30 : 70 K=11).....	51
Gambar 4. 7 <i>Confusion Matrix KNN Minmax</i> RFMT ( 20 : 80 K=9).....	51
Gambar 4. 8 <i>Confusion Matrix KNN Z Score</i> RMT ( 20 : 80 K=13).....	52
Gambar 4. 9 <i>Confusion Matrix KNN Minmax</i> RFMT ( 40 : 60 K=15).....	52



## DAFTAR TABEL

Tabel 2. 1 Tinjauan Pustaka.....	10
Tabel 2. 2 Tinjauan Pustaka (Lanjutan).....	11
Tabel 2. 3 Tinjauan Pustaka (Lanjutan).....	12
Tabel 2. 4 Perbedaan Minmax dan Z-score.....	29
Tabel 2. 5 <i>Confusion Matrix</i> .....	32
Tabel 4. 1 Atribut Pendonor.....	35
Tabel 4. 2 Data akan dinormalisasi.....	38
Tabel 4. 3 Nilai Mean.....	38
Tabel 4. 4 Nilai Standar Deviasi.....	39
Tabel 4. 5 Hasil <i>Z-Score Normalization</i> .....	40
Tabel 4. 6 Hasil <i>Minmax</i> .....	42
Tabel 4. 7 Contoh data Uji.....	43
Tabel 4. 8 Hasil Perhitungan Jarak.....	44
Tabel 4. 9 Contoh Hasil Perhitungan Jarak.....	46
Tabel 4. 10 Contoh Data Uji <i>Naive Bayes</i> .....	47
Tabel 4. 11 Contoh Data Latih <i>Naive Bayes</i> .....	48
Tabel 4. 12 Prosentase Akurasi <i>Naive Bayes</i> .....	53
Tabel 4. 13 Prosentase Akurasi <i>K-Nearest Neighbor</i> .....	53



**STUDI KOMPARASI METODE *NAIVE BAYES* DAN *K-NEAREST NEIGHBOR* (KNN) UNTUK MEMPREDIKSI PENDONOR DARAH POTENSIAL (STUDI KASUS: *BLOOD TRANSFUSION SERVICE CENTER*)**

**NUR FAIZAH**

**16650012**

**INTISARI**

Kurangnya stok pasokan kantong darah di Indonesia setiap tahun mengakibatkan angka kematian pada ibu hamil dan melahirkan cukup tinggi. Pentingnya pemenuhan stok kantong darah mampu meningkatkan pelayanan kesehatan dan menyelamatkan nyawa seseorang. Darah yang telah didonorkan tidak dapat digunakan setelah 42 hari, maka untuk memenuhi stok minimum kantong darah, maka pendonor darah potensial harus diketahui. Prediksi dapat dilakukan dengan menggunakan metode dalam data *mining* yang memiliki akurasi tinggi.

Prediksi dilakukan dengan menggunakan data *blood transfusion service centre* dari *UCI Machine Learning*. Data berjumlah 748 dengan 4 atribut kondisi pendonor untuk menentukan prediksi yang kemudian dievaluasi korelasi antar atribut dependen. Prediksi dilakukan dengan menggunakan *Naive Bayes* dan *K-Nearest Neighbor* yang kemudian dibandingkan nilai akurasi dari masing-masing metode untuk mendapatkan nilai tertinggi.

Berdasarkan Pengujian menggunakan *confusion matrix* algoritma *Naive Bayes* menghasilkan akurasi 76% untuk dataset RMT dengan menggunakan *preprocessing Z-Score Normalization* maupun *Minmax Scaler* sedangkan algoritma *K-Nearest Neighbor* (KNN) menghasilkan akurasi sebesar 79.3% pada perbandingan data uji 20% dan data latih 80% menggunakan dataset *Recency*, *Monetary* dan *Time*, dan *Z-Score Normalization* untuk *preprocessing* data. Sehingga dapat disimpulkan bahwa algoritma *K-Nearest Neighbor* (KNN) adalah algoritma yang paling relevan digunakan untuk memprediksi pendonor darah potensial.

Kata Kunci: Donor Darah, *Naive Bayes*, *K-Nearest Neighbor*, *Confusion Matrix*.

# **COMPARISON STUDY OF NAIVE BAYES AND K-NEAREST NEIGHBOR (KNN) TO PREDICT POTENTIAL BLOOD DONORS (CASE STUDY: BLOOD TRANSFUSION SERVICE CENTER)**

**NUR FAIZAH**

**16650012**

## **ABSTRACT**

The lack of supply of blood bags in Indonesia every year results in a high mortality rate in pregnant and childbirth mothers. The importance of fulfilling blood bag stocks can improve health care and save a person's life. Donated blood cannot be used after 42 days, so to meet the minimum stock of blood bags, then potential blood donors should be known. Predictions can be made using methods in data mining that have high accuracy.

Predictions are made using blood transfusion service center data from UCI Machine Learning. The data amounted to 748 with 4 donor condition attributes to determine predictions that were then evaluated correlations between dependent attributes. Predictions were made using Naive Bayes and K-Nearest Neighbor which then compared the accuracy values of each method to get the highest score.

Based on the test using the Naïve Bayes algorithm, the algorithm produces 76% accuracy for the RMT dataset using preprocessing Z-Score Normalization and Minmax Scaler while the K-Nearest Neighbor (KNN) algorithm produces an accuracy of 79.3% on a comparison of 20% test data and 80% training data using dataset Recency, Monetary and Time, and Z-Score Normalization for data preprocessing. So it can be concluded that the K-Nearest Neighbor (KNN) algorithm is the most relevant algorithm used to predict potential blood donors.

Keyword: Blood Donation, Naive Bayes, K-Nearest Neighbor, Confusion Matrix.



# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Menurut Daradjatun (2008) Donor darah adalah proses pengambilan darah dari seseorang secara sukarela untuk disimpan di bank darah sebagai stok darah yang kemudian digunakan untuk keperluan transfusi darah. Pentingnya pemenuhan kebutuhan darah mampu meningkatkan pelayanan kesehatan dan menyelamatkan nyawa seseorang (Kementerian Kesehatan Republik Indonesia, 2014). Ibu hamil dan melahirkan termasuk salah satu kelompok yang mengalami kerugian akibat kekurangan kantong darah. Angka kematian ibu hamil yang masih tinggi yaitu 359 per 100.000 kelahiran (SDKI, 2012). Kementerian Kesehatan menunjukkan bahwa penyebab kematian ibu melahirkan 35% akibat pendarahan.

Menurut *World Health Organization* (WHO) jumlah kebutuhan minimal darah adalah 2% dari jumlah penduduk. Menurut Bambang (2018) berdasarkan data dari WHO, kebutuhan darah di Indonesia per tahun mencapai 5,1 juta kantong darah, sementara yang terpenuhi hanya sekitar 4,2 juta kantong darah. Kurangnya ketersediaan pasokan kantong darah disebabkan karena kurangnya pendonor darah dan banyak pendonor darah yang ditolak karena tidak memenuhi syarat.

Berkembangnya teknologi, informasi dan komunikasi telah menghasilkan banyaknya konsep teknologi yang mampu menunjang kebutuhan informasi dan komunikasi. Salah satu konsep yang ditawarkan adalah data mining. Menurut Kusriani dan Taufiq (2009) Data mining adalah proses otomatis terhadap data

besar yang bertujuan untuk mendapatkan suatu pola. Data mining sangat bermanfaat dalam penggalian informasi dalam jumlah data yang besar. UCI machine learning merupakan salah satu laman yang menyediakan berbagai data besar (dataset) secara terbuka agar dapat dimanfaatkan oleh seseorang untuk menggali informasi dari data yang disediakan.

Di Indonesia seseorang yang bisa mendonorkan darahnya telah diatur dalam peraturan kesehatan, sedangkan karakteristik pendonor dia dapat mendonorkan darahnya kembali atau tidak dimasa akan datang dapat dibaca berdasarkan rekam medis selama menjadi pendonor. Dalam dataset UCI machine learning terdapat dataset *blood transfusion dataset* dengan variabel yang telah diteliti untuk meramalkan karakteristik pendonor darah. Variabel tersebut dapat diterapkan bagi peneliti khususnya dalam memprediksi pendonor darah potensial di Indonesia dengan studi data yang berbeda.

Sejauh ini telah banyak algoritma yang digunakan untuk memprediksi pendonor darah potensial. Dalam penelitian ini akan dilakukan studi komparasi metode *K-Nearest Neighbor* dan *Naive Bayes* untuk memprediksi pendonor darah potensial. *K-Nearest Neighbor* dan *Naive Bayes* merupakan metode pengklasifikasi yang terkenal dengan tingkat keakuratan yang baik.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah yang dapat disimpulkan adalah belum adanya studi komparasi penggunaan algoritma KNN dan *Naive Bayes* untuk memprediksi pendonor darah.

### 1.3 Tujuan Penelitian

Berdasarkan pada latar belakang dan rumusan masalah yang telah dibahas, maka tujuan penelitian ini adalah untuk mengetahui metode yang lebih akurat antara metode KNN dan *Naive Bayes* untuk prediksi pendonor darah.

### 1.4 Batasan Masalah

Berdasarkan rumusan masalah diatas, agar pembahasan tidak melebar dan terfokus pada tujuan yang diinginkan, maka batasan masalah dalam penelitian adalah sebagai berikut:

1. Penelitian dilakukan dengan menggunakan dataset *Blood Transfusion Service Centre Database* dari UCI Machine Learning.
2. Sistem hanya menghasilkan prediksi donor atau tidak.
3. Sistem hanya dibangun dengan bahasa pemrograman *python*.

### 1.5 Manfaat Penelitian

Hasil dari penelitian ini diharapkan mampu memberikan manfaat bagi peneliti dan pembaca.

1. Bagi Pembaca

Penelitian ini dapat membantu menambah wawasan dan memperluas pengetahuan terutama mampu mengetahui cara prediksi pendonor darah potensial menggunakan perbandingan metode *K-Nearest Neighbor* dan *Naive Bayes Classifier*.

2. Bagi peneliti lain

Penelitian ini sebagai bentuk sumbangan terhadap penelitian lainnya agar dapat mengembangkan penelitian lebih lanjut berkaitan dengan masalah perbandingan metode.

### 3. Bagi PMI

Penelitian ini dapat membantu PMI untuk membaca karakteristik pendonor di masa akan datang sehingga mampu memprediksi stok darah berikutnya mampu mencukupi target atau tidak.

#### 1.6 Keaslian Penelitian

Penelitian tentang prediksi pendonor darah potensial hingga saat ini sudah banyak dilakukan oleh peneliti sebelumnya. Namun, berdasarkan referensi dan tinjauan pustaka, penelitian yang diajukan sebagai Tugas Akhir S1 pada program studi Teknik Informatika Fakultas Sains Dan Teknologi UIN Sunan Kalijaga mengenai analisis perbandingan metode *K-Nearest Neighbor* dan Naive Bayes untuk memprediksi pendonor darah potensial dengan membandingkan seleksi fitur variabel berdasarkan korelasi sederhana dan membandingkan preprocessing z-score dan minmax belum pernah ditemukan oleh peneliti.

#### 1.7 Sistematika Penelitian

Laporan penelitian tugas akhir ini disusun secara sistematis dibagi dalam beberapa bab. Penyusunan laporan tugas akhir ini memiliki urutan yang diawali dari BAB I dan diakhiri BAB V :

**BAB I                    PENDAHULUAN**

Bab pertama ini berisikan keterangan mengenai latar belakang dilakukan penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, keaslian penelitian dan sistematika penulisan tugas akhir sebagai laporan hasil penelitian.

## **BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI**

Bab kedua pada penelitian ini berisikan penjelasan tentang tinjauan pustaka dan landasan teori yang berkaitan atau berhubungan dengan topik yang dibahas pada penelitian ini.

## **BAB III METODE PENELITIAN**

Bab ini berisikan tentang penjelasan mengenai langkah-langkah yang harus dilakukan untuk mencapai tujuan dan simpulan tugas akhir.

## **BAB IV HASIL DAN PEMBAHASAN**

Pada bab ini membahas analisis data dan hasil dari penelitian yang telah dilakukan.

## **BAB V PENUTUP**

Bab ini berisikan tentang kesimpulan yang dihasilkan serta saran yang akan diberikan berdasarkan

hasil yang telah dicapai sehingga dapat digunakan sebagai bahan pertimbangan bagi pihak-pihak yang berkepentingan serta kemungkinan untuk perkembangan pada penelitian-penelitian selanjutnya.



## **BAB V PENUTUP**

### **5.1 Kesimpulan**

Berdasarkan penelitian yang telah dilakukan didapatkan bahwa algoritma Naïve Bayes dan algoritma *K-Nearest Neighbor* (KNN) dapat digunakan untuk memprediksi pendonor darah pada studi kasus *Blood Transfusion Service Centre Database*.

Pengujian menggunakan *confusion matrix* terhadap kedua algoritma menyatakan bahwa algoritma Naïve Bayes menghasilkan akurasi 76% untuk dataset RMT dengan menggunakan *preprocessing Z-Score Normalization* maupun *Minmax Scaler* sedangkan algoritma *K-Nearest Neighbor* (KNN) menghasilkan akurasi sebesar 79.3% pada perbandingan data uji 20% dan data latih 80% menggunakan dataset *Recency, Monetary* dan *Time*, dan *Z-Score Normalization* untuk *preprocessing* data. Sehingga dapat disimpulkan bahwa algoritma *K-Nearest Neighbor* (KNN) adalah algoritma yang paling relevan digunakan untuk memprediksi pendonor darah potensial.

### **5.2 Saran**

Pada penelitian ini tentu masih banyak kekurangan, oleh karena itu penulis memberikan saran yang bias dijadikan perbaikan untuk pengembangan penelitian selanjutnya. Saran-saran tersebut meliputi:

1. Penelitian prediksi donor darah sebaiknya lebih melibatkan banyak data dengan perbandingan kelas variable dependen yang tidak jauh berbeda. Sehingga nantinya dapat menghasilkan akurasi yang lebih akurat.

2. Penelitian selanjutnya dapat menggunakan atribut, metode dan preprocessing yang lainnya atau mengkombinasikan beberapa metode.
3. Penelitian selanjutnya bisa membuat sistem prediksi data pendonor darah secara otomatis.





## DAFTAR PUSTAKA

- Bambang. 2018. *Kebutuhan Darah di Indonesia Belum Terpenuhi*  
<https://mediaindonesia.com/read/detail/242935-kebutuhan-darah-di-indonesia-belum-terpenuhi>
- Bustami, (2013). Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi, *TECHSI : Jurnal Penelitian Teknik Informatika*.
- Dania, A. N., Hidayah, H. K., & Nurul, C. (2019). *Perbandingan Normalisasi Data Untuk Klasifikasi Wine Menggunakan Algoritma K-NN*. *Jurnal of Computer Engineering System and Science* 4(1), 78–82.
- Daradjatun. (2008). *Pedoman Pelayanan Tranfusi Darah*. Jakarta: UTD PMI Pusat
- Darwiche, M., Feuilo, M., Bousaleh, G., & Schang D. (2010). *Prediction of blood transfusion donation*. *IEEE*, 978.
- Davies, dan Paul Beynon. 2004. *Database Systems and Third Edition*. New York: Palgrave Macmillan
- Elvianti. (2015). *Penerapan Metode Modified K-Nearest Neighbor (MK-NN) Untuk Klasifikasi Penderita Liver*. Riau: UIN Sultan Syarif Kasmin
- Eska, J. (2016). *Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5*. *Jurnal Teknologi dan Sistem Infomasi*, 9-13
- Fa'rifah, R., & Purhadi. (2012). Analisis Survival Fakto-Faktor yang Mempengaruhi Laju Kesembuhan Pasien Penderita Demam Berdaha Dengue (DBD) di SU Haji Surabaya dengan Regesi Cox. *Jurnal Sains Dan Seni Its*, 271-276.
- Fais A, S. N., Aditya D, M., & Mulya I, S. (2015). *Klasifikasi calon Pendoron Darah dengan Metode Naïve Bayes Classifier*. Malang: Universitas Brawijaya.
- Gorunescu, Florin. (2011). *Data Mining: Concepts, Models, and Techniques*. Verlag Berlin Heidelberg: Springer.
- Han, Jiawie dan Kamber. (2006). *Data Mining : Concept and Texhniques, 2<sup>nd</sup> edition*. New York: Morgan Kaufman.
- Kementerian Kesehatan Republik Indonesia. (24 Juni 2014). *Darah Untuk MenyelamatkanIbu*.  
<https://www.kemkes.go.id/article/view/201406250001/darah-untuk-menyelamatkan-ibu.html>

- Kesumadewi, Sri. *Klasifikasi Status Gizi Menggunakan Naïve Bayes Classification*
- Kusrini, (2008). *Konsep dan Aplikasi Sistem Pakar*. Yogyakarta: Andi.
- Kusrini dan Taufiq, L. E. (2009). *Algoritma Data Mining*. Yogyakarta: Andi.
- Mohamad Efendi Lasulika. (2019). *Komparasi Naïve Bayes , Support Vector Machine Dan K- Nearest Neighbor Untuk Mengetahui Akurasi Tertinggi*. Gorontalo: Universitas Ichsan Gorontalo.
- Mohamad Azmar Maricar, dkk. (2019). *Perbandingan Akurasi Naïve Bayes dan K-Nearest Neighbor pada Klasifikasi untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali*. *Jurnal Institut Teknologi dan Bisnis Bali*, 16–22.
- Mustafa, M.S., Ramadhan, M., & Thenata, A. P. (2017). *Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier*. Makasar: STIMIK Dipanegara.
- Ningkrat, dkk. (2018). *Klasifikasi Penentuan Calon Pendorong Darah dengan Menggunakan Metode K-Nearest Neighbor (Studi Kasus PMI Kab Gresik)*. Gresik: Universitas Muhammadiyah Gresik.
- Astrid Noviriandini. (2019). *Prediksi Penyakit Liver dengan Menggunakan Metode Naive Bayes dan K-Nearest Neighbor (KNN)*. *Seminar Nasional Rekayasa dan Teknologi*, 75-80.
- Nugroho, E. B., (2018). *Klasifikasi Pendorong Darah Menggunakan Metode Support Vector Machine*. *Jurnal Universitas Brawijaya* 2 (10), 3860–3864.
- Shadab Adam Pattekari, Asma Parveen. (2012). *Prediction System for Heart Disease Using Naive Bayes*, *International Journal of Advanced Computer and Mathematical Sciences*, ISSN 2230-9624, Vol. 3, Issue 3.
- Permenkes RI. 2015. *Standar Pelayanan Transfusi Darah*. Jakarta : Depkes.
- Pranoto, A. S. (2018) *Analisis Perbandingan Algoritma K-Nearest Neighbor Bank Tabungan Negara Cabang Puri Indah*. Jakarta: Sekolah Tinggi Manajemen Informatika dan Komputer Eresha.
- Prasetyo. (2012) *Data Mining – Konsep dan Aplikasi Menggunakan Matlab*, edisi 1, Yogyakarta: Andi.
- Prasetyo. (2014). *Data Mining- Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi.

- Pusat Data dan Informasi Kementerian Kesehatan RI, (2014), *Situasi Donor Darah di Indonesia*. Jakarta Selatan, Pusat Data dan Informasi Kementerian Kesehatan RI.
- Maulana Aditya Rahman, dkk. (2018). *Komparasi Metode Data Mining K-Nearest Neighbor Dengan Naïve Bayes Untuk Klasifikasi Kualitas Air Bersih (Studi Kasus PDAM Tirta Kencana Kabupaten Jomban)*. *Jurnal Universitas Brawijaya*, 2(12).
- Rifai, M. F., Jatnika, H., & Valentino, B. (2019). *Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS)*. *Jurnal Pengkajian dan Penerapan Teknik Informatika* 12(2), 131–144.
- Rodiyansah, S. F. (2016). Naïve Bayes Classification untuk Penentuan Kelayakan Donor Darah, *Jurnal Universitas Majalengka*, 156-159.
- Rohmawan, E. P. (2019). Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree dan Artificial Neural Network. *Jurnal Ilmiah Matrik Universitas Bina Dharma*, 23-30.
- Santoso, B. (2007). *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- SDKI. (2012). *Survei demografi dan kesehatan Indonesia*. Jakarta
- Schuerer dan Maufrais (2010). *Introduction to Programming using Python*. Boston: Pearson, hal. 1–242. ISBN: 0132747189.
- Silalahi, D. K., Murfi, H., & Satria, Y. (2017). Studi Perbandingan Pemilihan Fitur untuk Support Vector Machine pada Klasifikasi Penilaian Risiko Kredit. *Jurnal EduMatSains* 1(2), 119–136.
- Simanjutak, Tri, H., Mahmudy, W., F., & Sutrisno. *Implementasi Modified K-Nearest Neighbor dengan Otomatisasi Nilai K pada Pengklasifikasian Penyakit Tanaman Kedelai*. Malang: Universitas Brawijaya.
- Suliztia, M. L., & Fauzan, A. (2020). Comparing Naive Bayes , K-Nearest Neighbor , And Neural Network Classification Methods Of Seat Load Factor In Lombok Outbound Flights. *Jurnal Matematika, Statistika dan Komputasi* 16(2), 187–198.
- Sulastris, dkk. (2019). *Analisis Perbandingan Klasifikasi Prediksi Penyakit Hepatitis Dengan Menggunakan Algoritma K-Nearest Neighbor, Naive Bayes dan Neural Network*. Semarang: Universitas Stikubank.

- Suyanto. (2020). *Data Mining Untuk Klasifikasi dan Klusterisasi Data*. Bandung: Informatika.
- Tan, dkk. (2005). *Introduction to Data Mining, 1 st Ed.*, Addison-Wesley, Boston.
- MacLennan, J., dkk. (2009). *Data Mining with Microsoft SQL Server 2008*. USA: WileyPublishing,Inc.
- Utomo, D. P. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidama*, 437–444.
- Wahono, H., & Riana, D. (2020). Prediksi Calon Pendorong Darah Potensial Dengan Algoritma Naïve Bayes , K- Nearest Neighbors dan Decision Tree C4.5. *Jurnal Riset Komputer* 7(1), 7–14.
- Paulus Wicaksana. (2015). *Perbandingan Algoritma K-Nearest Neighbors Dan Naive Bayes Untuk Studi Data “Wisconsin Diagnosis Breast Cancer”*. Yogyakarta: Univesitas Sanata Dharma.
- Wahyu Eko Susanto, dkk (2016). *Komparasi Algoritma Neural Network, K-Nearest Neighbor dan Naive Bayes untuk Memprediksi Pendorong Darah Potensial*. *Jurnal Speed*, Vol 8 No 3.
- Witten, I. H., & Frank, E. (2005). *Data Mining : Practical Machine Learning Tools and Techniques Second Edition*. Morgan Kauffman : San Francisco
- Yunus, M., Dahlan, H. S., & Santoso, P. B. (2014). SPK Pemilihan Calon Pendorong Darah Potensial Fuzzi Tahani, *Jurnal EECCIS*, 47-54.

## LAMPIRAN

Dataset:

R,F,M,T,"status"	11 ,2,500,26 ,0
2 ,50,12500,98 ,1	16 ,5,1250,40 ,1
0 ,13,3250,28 ,1	4 ,2,500,51 ,1
1 ,16,4000,35 ,1	14 ,3,750,31 ,0
2 ,20,5000,45 ,1	4 ,2,500,52 ,0
1 ,24,6000,77 ,0	9 ,4,1000,65 ,0
4 ,4,1000,4 ,0	14 ,4,1000,40 ,0
2 ,7,1750,14 ,1	11 ,3,750,40 ,1
1 ,12,3000,35 ,0	14 ,5,1250,50 ,0
2 ,9,2250,22 ,1	14 ,1,250,14 ,0
5 ,46,11500,98 ,1	14 ,1,250,14 ,0
4 ,23,5750,58 ,0	14 ,1,250,14 ,0
0 ,3,750,4 ,0	14 ,1,250,14 ,0
2 ,10,2500,28 ,1	14 ,1,250,14 ,0
1 ,13,3250,47 ,0	14 ,1,250,14 ,0
2 ,6,1500,15 ,1	14 ,1,250,14 ,0
2 ,5,1250,11 ,1	14 ,1,250,14 ,0
2 ,14,3500,48 ,1	14 ,7,1750,72 ,0
2 ,15,3750,49 ,1	14 ,1,250,14 ,0
2 ,6,1500,15 ,1	14 ,1,250,14 ,0
2 ,3,750,4 ,1	9 ,3,750,52 ,0
2 ,3,750,4 ,1	14 ,7,1750,73 ,0
4 ,11,2750,28 ,0	11 ,4,1000,58 ,0
2 ,6,1500,16 ,1	11 ,4,1000,59 ,0
2 ,6,1500,16 ,1	4 ,2,500,59 ,0
9 ,9,2250,16 ,0	11 ,4,1000,61 ,0
4 ,14,3500,40 ,0	16 ,4,1000,40 ,0
4 ,6,1500,14 ,0	16 ,10,2500,89 ,0
4 ,12,3000,34 ,1	21 ,2,500,21 ,1
4 ,5,1250,11 ,1	21 ,3,750,26 ,0
4 ,8,2000,21 ,0	16 ,8,2000,76 ,0
1 ,14,3500,58 ,0	21 ,3,750,26 ,1
4 ,10,2500,28 ,1	18 ,2,500,23 ,0
4 ,10,2500,28 ,1	23 ,5,1250,33 ,0
4 ,9,2250,26 ,1	23 ,8,2000,46 ,0
2 ,16,4000,64 ,0	16 ,3,750,34 ,0
2 ,8,2000,28 ,1	14 ,5,1250,64 ,0
2 ,12,3000,47 ,1	14 ,3,750,41 ,0
4 ,6,1500,16 ,1	16 ,1,250,16 ,0
2 ,14,3500,57 ,1	16 ,1,250,16 ,0
4 ,7,1750,22 ,1	16 ,1,250,16 ,0
2 ,13,3250,53 ,1	16 ,1,250,16 ,0

2,5,1250,16,0	16,1,250,16,0
2,5,1250,16,1	16,1,250,16,0
2,5,1250,16,0	16,1,250,16,0
4,20,5000,69,1	16,4,1000,45,0
4,9,2250,28,1	16,1,250,16,0
2,9,2250,36,0	16,1,250,16,0
2,2,500,2,0	16,1,250,16,0
2,2,500,2,0	16,1,250,16,0
2,2,500,2,0	16,1,250,16,0
2,11,2750,46,0	16,2,500,26,0
2,11,2750,46,1	21,2,500,23,0
2,6,1500,22,0	16,2,500,27,0
2,12,3000,52,0	21,2,500,23,0
4,5,1250,14,1	21,2,500,23,0
4,19,4750,69,1	14,4,1000,57,0
4,8,2000,26,1	16,5,1250,60,0
2,7,1750,28,1	23,2,500,23,0
2,16,4000,81,0	14,5,1250,74,0
3,6,1500,21,0	23,3,750,28,0
2,7,1750,29,0	16,3,750,40,0
2,8,2000,35,1	9,2,500,52,0
2,10,2500,49,0	9,2,500,52,0
4,5,1250,16,1	16,7,1750,87,1
2,3,750,9,1	14,4,1000,64,0
3,16,4000,74,0	14,2,500,35,0
2,4,1000,14,1	16,7,1750,93,0
0,2,500,4,0	21,2,500,25,0
4,7,1750,25,0	14,3,750,52,0
1,9,2250,51,0	23,14,3500,93,0
2,4,1000,16,0	18,8,2000,95,0
2,4,1000,16,0	16,3,750,46,0
4,17,4250,71,1	11,3,750,76,0
2,2,500,4,0	11,2,500,52,0
2,2,500,4,1	11,3,750,76,0
2,2,500,4,1	23,12,3000,86,0
2,4,1000,16,1	21,3,750,35,0
2,2,500,4,0	23,2,500,26,0
2,2,500,4,0	23,2,500,26,0
2,2,500,4,0	23,8,2000,64,0
4,6,1500,23,1	16,3,750,50,0
2,4,1000,16,0	23,3,750,33,0
2,4,1000,16,0	21,3,750,38,0
2,4,1000,16,0	23,2,500,28,0
2,6,1500,28,1	21,1,250,21,0
2,6,1500,28,0	21,1,250,21,0
4,2,500,4,0	21,1,250,21,0

4 ,2,500,4 ,0	21 ,1,250,21 ,0
4 ,2,500,4 ,0	21 ,1,250,21 ,0
2 ,7,1750,35 ,1	21 ,1,250,21 ,0
4 ,2,500,4 ,1	21 ,1,250,21 ,0
4 ,2,500,4 ,0	21 ,1,250,21 ,0
4 ,2,500,4 ,0	21 ,1,250,21 ,0
4 ,2,500,4 ,0	21 ,1,250,21 ,1
12 ,11,2750,23 ,0	21 ,1,250,21 ,0
4 ,7,1750,28 ,0	21 ,1,250,21 ,0
3 ,17,4250,86 ,0	21 ,5,1250,60 ,0
4 ,9,2250,38 ,1	23 ,4,1000,45 ,0
4 ,4,1000,14 ,1	21 ,4,1000,52 ,0
5 ,7,1750,26 ,1	22 ,1,250,22 ,1
4 ,8,2000,34 ,1	11 ,2,500,70 ,0
2 ,13,3250,76 ,1	23 ,5,1250,58 ,0
4 ,9,2250,40 ,0	23 ,3,750,40 ,0
2 ,5,1250,26 ,0	23 ,3,750,41 ,0
2 ,5,1250,26 ,0	14 ,3,750,83 ,0
6 ,17,4250,70 ,0	21 ,2,500,35 ,0
0 ,8,2000,59 ,0	26 ,5,1250,49 ,1
3 ,5,1250,26 ,0	23 ,6,1500,70 ,0
2 ,3,750,14 ,0	23 ,1,250,23 ,0
2 ,10,2500,64 ,0	23 ,1,250,23 ,0
4 ,5,1250,23 ,1	23 ,1,250,23 ,0
4 ,9,2250,46 ,0	23 ,1,250,23 ,0
4 ,5,1250,23 ,0	23 ,1,250,23 ,0
4 ,8,2000,40 ,1	23 ,1,250,23 ,0
2 ,12,3000,82 ,0	23 ,1,250,23 ,0
11 ,24,6000,64 ,0	23 ,1,250,23 ,0
2 ,7,1750,46 ,1	23 ,4,1000,53 ,0
4 ,11,2750,61 ,0	21 ,6,1500,86 ,0
1 ,7,1750,57 ,0	23 ,3,750,48 ,0
2 ,11,2750,79 ,1	21 ,2,500,41 ,0
2 ,3,750,16 ,1	21 ,3,750,64 ,0
4 ,5,1250,26 ,1	16 ,2,500,70 ,0
2 ,6,1500,41 ,1	21 ,3,750,70 ,0
2 ,5,1250,33 ,1	23 ,4,1000,87 ,0
2 ,4,1000,26 ,0	23 ,3,750,89 ,0
2 ,5,1250,34 ,0	23 ,2,500,87 ,0
4 ,8,2000,46 ,1	35 ,3,750,64 ,0
2 ,4,1000,26 ,0	38 ,1,250,38 ,0
4 ,8,2000,48 ,1	38 ,1,250,38 ,0
2 ,2,500,10 ,1	40 ,1,250,40 ,0
4 ,5,1250,28 ,0	74 ,1,250,74 ,0
2 ,12,3000,95 ,0	2 ,43,10750,86 ,1
2 ,2,500,10 ,0	6 ,22,5500,28 ,1

4,6,1500,35,0	2,34,8500,77,1
2,11,2750,88,0	2,44,11000,98,0
2,3,750,19,0	0,26,6500,76,1
2,5,1250,37,0	2,41,10250,98,1
2,12,3000,98,0	3,21,5250,42,1
9,5,1250,19,0	2,11,2750,23,0
2,2,500,11,0	2,21,5250,52,1
2,9,2250,74,0	2,13,3250,32,1
5,14,3500,86,0	4,4,1000,4,1
4,3,750,16,0	2,11,2750,26,0
4,3,750,16,0	2,11,2750,28,0
4,2,500,9,1	3,14,3500,35,0
4,3,750,16,1	4,16,4000,38,1
6,3,750,14,0	4,6,1500,14,0
2,2,500,11,0	3,5,1250,12,1
2,2,500,11,1	4,33,8250,98,1
2,2,500,11,0	3,10,2500,33,1
2,7,1750,58,1	4,10,2500,28,1
4,6,1500,39,0	2,11,2750,40,1
4,11,2750,78,0	2,11,2750,41,1
2,1,250,2,1	4,13,3250,39,1
2,1,250,2,0	1,10,2500,43,1
2,1,250,2,0	4,9,2250,28,0
2,1,250,2,0	2,4,1000,11,0
2,1,250,2,0	2,5,1250,16,1
2,1,250,2,0	2,15,3750,64,0
2,1,250,2,0	5,24,6000,79,0
2,1,250,2,0	2,6,1500,22,1
2,1,250,2,0	4,5,1250,16,1
2,1,250,2,0	2,4,1000,14,1
2,1,250,2,1	4,8,2000,28,0
2,1,250,2,1	2,4,1000,14,0
2,1,250,2,1	2,6,1500,26,0
2,1,250,2,0	4,5,1250,16,1
2,1,250,2,0	2,7,1750,32,1
2,1,250,2,0	2,6,1500,26,1
2,1,250,2,0	2,8,2000,38,1
2,1,250,2,0	2,2,500,4,1
2,1,250,2,0	2,6,1500,28,1
2,1,250,2,0	2,10,2500,52,0
2,1,250,2,0	4,16,4000,70,1
2,1,250,2,0	4,2,500,4,1
11,10,2500,35,0	1,14,3500,95,0
11,4,1000,16,1	4,2,500,4,1
4,5,1250,33,1	7,14,3500,48,0
4,6,1500,41,1	2,3,750,11,0



2,3,750,22,0	2,12,3000,70,1
4,4,1000,26,1	4,7,1750,32,1
10,4,1000,16,0	4,4,1000,16,0
2,4,1000,35,0	2,6,1500,35,1
4,12,3000,88,0	4,6,1500,28,1
13,8,2000,26,0	2,3,750,14,0
11,9,2250,33,0	2,4,1000,23,0
4,5,1250,34,0	4,4,1000,18,0
4,4,1000,26,0	5,6,1500,28,0
8,15,3750,77,0	4,6,1500,30,0
4,5,1250,35,1	14,5,1250,14,0
4,7,1750,52,0	3,8,2000,50,0
4,7,1750,52,0	4,11,2750,64,1
2,4,1000,35,0	4,9,2250,52,0
11,11,2750,42,0	4,16,4000,98,1
2,2,500,14,0	7,10,2500,47,0
2,5,1250,47,1	4,14,3500,86,0
9,8,2000,38,1	2,9,2250,75,0
4,6,1500,47,0	4,6,1500,35,0
11,7,1750,29,0	4,9,2250,55,0
9,9,2250,45,0	4,6,1500,35,1
4,6,1500,52,0	2,6,1500,45,0
4,7,1750,58,0	2,6,1500,47,0
6,2,500,11,1	4,2,500,9,0
4,7,1750,58,0	2,2,500,11,1
11,9,2250,38,0	2,2,500,11,0
11,6,1500,26,0	2,2,500,11,1
2,2,500,16,0	4,6,1500,38,1
2,7,1750,76,0	3,4,1000,29,1
11,6,1500,27,0	9,9,2250,38,0
11,3,750,14,0	11,5,1250,18,0
4,1,250,4,0	2,3,750,21,0
4,1,250,4,0	2,1,250,2,0
4,1,250,4,0	2,1,250,2,1
4,1,250,4,0	2,1,250,2,0
4,1,250,4,0	2,1,250,2,0
4,1,250,4,0	2,1,250,2,0
4,1,250,4,0	2,1,250,2,1
4,1,250,4,0	2,1,250,2,0
4,1,250,4,0	2,1,250,2,0
4,1,250,4,0	2,1,250,2,0
4,1,250,4,1	2,1,250,2,0
4,1,250,4,1	11,11,2750,38,0
4,1,250,4,0	2,3,750,22,0
4,1,250,4,1	9,11,2750,49,1

4 ,1,250,4 ,1	5 ,11,2750,75 ,0
4 ,1,250,4 ,0	3 ,5,1250,38 ,0
4 ,3,750,24 ,0	3 ,1,250,3 ,1
4 ,1,250,4 ,0	4 ,6,1500,43 ,0
4 ,1,250,4 ,0	2 ,3,750,24 ,0
4 ,1,250,4 ,0	12 ,11,2750,39 ,0
4 ,1,250,4 ,1	2 ,2,500,14 ,0
4 ,1,250,4 ,0	4 ,6,1500,46 ,0
10 ,8,2000,39 ,0	9 ,3,750,14 ,0
14 ,7,1750,26 ,0	14 ,8,2000,26 ,0
8 ,10,2500,63 ,0	4 ,2,500,13 ,0
11 ,3,750,15 ,0	4 ,11,2750,95 ,0
4 ,2,500,14 ,0	2 ,7,1750,77 ,0
2 ,4,1000,43 ,0	2 ,7,1750,77 ,0
8 ,9,2250,58 ,0	4 ,1,250,4 ,0
8 ,8,2000,52 ,1	4 ,1,250,4 ,0
11 ,22,5500,98 ,0	4 ,1,250,4 ,0
4 ,3,750,25 ,1	4 ,1,250,4 ,0
11 ,17,4250,79 ,1	4 ,1,250,4 ,1
9 ,2,500,11 ,0	4 ,1,250,4 ,0
4 ,5,1250,46 ,0	4 ,1,250,4 ,0
11 ,12,3000,58 ,0	4 ,1,250,4 ,0
7 ,12,3000,86 ,0	4 ,1,250,4 ,0
11 ,2,500,11 ,0	4 ,1,250,4 ,0
11 ,2,500,11 ,0	4 ,1,250,4 ,1
11 ,2,500,11 ,0	4 ,1,250,4 ,0
2 ,6,1500,75 ,0	4 ,7,1750,62 ,0
11 ,8,2000,41 ,1	4 ,1,250,4 ,0
11 ,3,750,16 ,1	4 ,4,1000,34 ,1
12 ,13,3250,59 ,0	11 ,6,1500,28 ,0
2 ,3,750,35 ,0	13 ,3,750,14 ,1
16 ,8,2000,28 ,0	7 ,5,1250,35 ,0
11 ,7,1750,37 ,0	9 ,9,2250,54 ,0
4 ,3,750,28 ,0	11 ,2,500,11 ,0
12 ,12,3000,58 ,0	2 ,5,1250,63 ,0
4 ,4,1000,41 ,0	7 ,11,2750,89 ,0
11 ,14,3500,73 ,1	8 ,9,2250,64 ,0
2 ,2,500,23 ,0	2 ,2,500,22 ,0
2 ,3,750,38 ,1	6 ,3,750,26 ,0
4 ,5,1250,58 ,0	12 ,15,3750,71 ,0
4 ,4,1000,43 ,1	13 ,3,750,16 ,0
3 ,2,500,23 ,0	11 ,16,4000,89 ,0
11 ,8,2000,46 ,0	4 ,5,1250,58 ,0
4 ,7,1750,82 ,0	14 ,7,1750,35 ,0
13 ,4,1000,21 ,0	11 ,4,1000,27 ,0
16 ,11,2750,40 ,0	7 ,9,2250,89 ,1

16,7,1750,28,0	11,8,2000,52,1
7,2,500,16,0	7,5,1250,52,0
4,5,1250,58,0	11,6,1500,41,0
4,5,1250,58,0	10,5,1250,38,0
4,4,1000,46,0	14,2,500,14,1
14,13,3250,57,0	14,2,500,14,0
4,3,750,34,0	14,2,500,14,0
14,18,4500,78,0	2,2,500,33,0
11,8,2000,48,0	11,3,750,23,0
14,16,4000,70,0	14,8,2000,46,0
14,4,1000,22,1	9,1,250,9,0
14,5,1250,26,0	16,5,1250,27,0
8,2,500,16,0	14,4,1000,26,0
11,5,1250,33,0	4,2,500,30,0
11,2,500,14,0	14,3,750,21,0
4,2,500,23,0	16,16,4000,77,0
9,2,500,16,1	4,2,500,31,0
14,5,1250,28,1	14,8,2000,50,0
14,3,750,19,1	11,3,750,26,0
14,4,1000,23,1	14,7,1750,45,0
16,12,3000,50,0	15,5,1250,33,0
11,4,1000,28,0	16,2,500,16,0
11,5,1250,35,0	16,3,750,21,0
11,5,1250,35,0	11,8,2000,72,0
2,4,1000,70,0	11,1,250,11,0
14,5,1250,28,0	11,1,250,11,0
14,2,500,14,0	11,1,250,11,0
14,2,500,14,0	11,1,250,11,1
14,2,500,14,0	11,1,250,11,0
14,2,500,14,0	2,3,750,75,1
14,2,500,14,0	2,3,750,77,0
14,2,500,14,0	16,4,1000,28,0
2,3,750,52,0	16,15,3750,87,0
14,6,1500,34,0	16,14,3500,83,0
11,5,1250,37,1	16,10,2500,62,0
4,5,1250,74,0	16,3,750,23,0
11,3,750,23,0	14,3,750,26,0
16,4,1000,23,0	23,19,4750,62,0
16,3,750,19,0	11,7,1750,75,0
11,5,1250,38,0	14,3,750,28,0
11,2,500,16,0	20,14,3500,69,1
12,9,2250,60,0	4,2,500,46,0
9,1,250,9,0	11,2,500,25,0
9,1,250,9,0	11,3,750,37,0
4,2,500,29,0	16,4,1000,33,0
11,2,500,17,0	21,7,1750,38,0

14 ,4,1000,26 ,0	13 ,7,1750,76 ,0
11 ,9,2250,72 ,1	16 ,6,1500,50 ,0
11 ,5,1250,41 ,0	14 ,3,750,33 ,0
15 ,16,4000,82 ,0	14 ,1,250,14 ,0
9 ,5,1250,51 ,1	14 ,1,250,14 ,0
11 ,4,1000,34 ,0	14 ,1,250,14 ,0
14 ,8,2000,50 ,1	14 ,1,250,14 ,0
16 ,7,1750,38 ,0	14 ,1,250,14 ,0
14 ,2,500,16 ,0	14 ,1,250,14 ,0
2 ,2,500,41 ,0	17 ,7,1750,58 ,1
14 ,16,4000,98 ,0	14 ,3,750,35 ,0
14 ,4,1000,28 ,1	14 ,3,750,35 ,0
16 ,7,1750,39 ,0	16 ,7,1750,64 ,0
14 ,7,1750,47 ,0	21 ,2,500,21 ,0
16 ,6,1500,35 ,0	16 ,3,750,35 ,0
16 ,6,1500,35 ,1	16 ,1,250,16 ,0
11 ,7,1750,62 ,1	16 ,1,250,16 ,0
16 ,2,500,16 ,0	16 ,1,250,16 ,0
16 ,3,750,21 ,1	16 ,1,250,16 ,0
11 ,3,750,28 ,0	16 ,1,250,16 ,0
11 ,7,1750,64 ,0	14 ,2,500,29 ,0
11 ,1,250,11 ,1	11 ,4,1000,74 ,0
9 ,3,750,34 ,0	11 ,2,500,38 ,1
14 ,4,1000,30 ,0	21 ,6,1500,48 ,0
23 ,38,9500,98 ,0	23 ,2,500,23 ,0
11 ,6,1500,58 ,0	23 ,6,1500,45 ,0
11 ,1,250,11 ,0	14 ,2,500,35 ,1
11 ,1,250,11 ,0	16 ,6,1500,81 ,0
11 ,1,250,11 ,0	16 ,4,1000,58 ,0
11 ,1,250,11 ,0	16 ,5,1250,71 ,0
11 ,1,250,11 ,0	21 ,2,500,26 ,0
11 ,1,250,11 ,0	21 ,3,750,35 ,0
11 ,1,250,11 ,0	21 ,3,750,35 ,0
11 ,1,250,11 ,0	23 ,8,2000,69 ,0
11 ,2,500,21 ,0	21 ,3,750,38 ,0
11 ,5,1250,50 ,0	23 ,3,750,35 ,0
11 ,2,500,21 ,0	21 ,3,750,40 ,0
16 ,4,1000,28 ,0	23 ,2,500,28 ,0
4 ,2,500,41 ,0	21 ,1,250,21 ,0
16 ,6,1500,40 ,0	21 ,1,250,21 ,0
14 ,3,750,26 ,0	25 ,6,1500,50 ,0
9 ,2,500,26 ,0	21 ,1,250,21 ,0
21 ,16,4000,64 ,0	21 ,1,250,21 ,0
14 ,6,1500,51 ,0	23 ,3,750,39 ,0
11 ,2,500,24 ,0	21 ,2,500,33 ,0
4 ,3,750,71 ,0	14 ,3,750,79 ,0

21 ,13,3250,57 ,0	23 ,1,250,23 ,1
11 ,6,1500,71 ,0	23 ,1,250,23 ,0
14 ,2,500,21 ,1	23 ,1,250,23 ,0
23 ,15,3750,57 ,0	23 ,1,250,23 ,0
14 ,4,1000,38 ,0	23 ,1,250,23 ,0
16 ,3,750,86 ,0	23 ,1,250,23 ,0
23 ,2,500,38 ,0	23 ,1,250,23 ,0
21 ,2,500,52 ,0	23 ,4,1000,52 ,0
23 ,3,750,62 ,0	23 ,1,250,23 ,0
39 ,1,250,39 ,0	23 ,7,1750,88 ,0
72 ,1,250,72 ,0	

Source Code KNN Minmax :

```
import pandas as pd, matplotlib.pyplot as plt, numpy as np, seaborn as sns
sns.set(style="ticks", color_codes=True)
from pandas.plotting import scatter_matrix
from sklearn import model_selection, tree
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.preprocessing import MinMaxScaler
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import cross_val_score
data = pd.read_csv('rfmtc.data')
data['status'] = data['status'].astype('category')
df = data.copy()
X = df[['Recency', 'Frequency', 'Monetary', 'Time']]
Y = df['status']
X_train, X_test, Y_train, Y_test = model_selection.train_test_split(X, Y,
test_size=0.2, random_state=123)
print(X_train.shape, len(Y_test))
```

```
sc = MinMaxScaler()

sc.fit(X_train)

X_train = sc.transform(X_train)

X_test = sc.transform(X_test)

print(X_train)

knn = KNeighborsClassifier(n_neighbors=11)

knn.fit(X_train, Y_train)

y_prediction = knn.predict(X_test)

print('Akurasi = ', accuracy_score(Y_test, y_prediction))

print(confusion_matrix(Y_test, y_prediction))

print(classification_report(Y_test, y_prediction))

#melihat korelasi

corr = data.corr()

# plot the heatmap

sns.heatmap(corr, annot=True)

plt.show()

df = data.copy()

# Separate Data

#Variabel RMT

X = df[['Recency','Monetary','Time']]

Y = df['status']

X_train, X_test, Y_train, Y_test = model_selection.train_test_split(X, Y,
test_size=0.2, random_state=123)
```

```

sc = MinMaxScaler()

sc.fit(X_train)

X_train = sc.transform(X_train)

X_test = sc.transform(X_test)

print(X_train)

knn = KNeighborsClassifier(n_neighbors=11)

knn = knn.fit(X_train, Y_train)

knn_prediction = knn.predict(X_test)

print('Akurasi = ', accuracy_score(Y_test, knn_prediction))

print(confusion_matrix(Y_test, knn_prediction))

print(classification_report(Y_test, knn_prediction))

#Source code Naive Bayes Z-score.

import pandas as pd, matplotlib.pyplot as plt, numpy as np, seaborn as sns

sns.set(style="ticks", color_codes=True)

from pandas.plotting import scatter_matrix

from sklearn import model_selection, tree

from sklearn.metrics import classification_report

from sklearn.metrics import confusion_matrix

from sklearn.metrics import accuracy_score

from sklearn.naive_bayes import GaussianNB

from sklearn.model_selection import cross_val_score

from sklearn.preprocessing import StandardScaler

data = pd.read_csv('rfmtc.data')

df = data.copy()

```

```

#semua variabel

# Separate Data

X = df[['Recency','Frequency','Monetary','Time']]

Y = df['status']

X_train, X_test, Y_train, Y_test = model_selection.train_test_split(X, Y,
test_size=0.4, random_state=123)

print(X_train.shape, len(Y_test))

sc = StandardScaler()

X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

gnb = GaussianNB()

nbc = gnb.fit(X_train, Y_train)

nbc_prediction = nbc.predict(X_test)

f, ax = plt.subplots(figsize=(5,3))

sns.heatmap(confusion_matrix(Y_test, nbc_prediction), annot=True, fmt=".0f",
ax=ax)

plt.show()
print('Akurasi = ', accuracy_score(Y_test, nbc_prediction))
print(classification_report(Y_test, nbc_prediction))

#feature seleksi

corr = data.corr()

# plot the heatmap

sns.heatmap(corr,

            xticklabels=corr.columns,

            yticklabels=corr.columns)

```



```
plt.show()

#Variabel RMT

df = data.copy()

# Separate Data

X = df[['Recency','Monetary','Time']]

Y = df['status']

#seed = 9

#validation_size = 0.4

X_train, X_test, Y_train, Y_test = model_selection.train_test_split(X, Y,
test_size=0.4, random_state=123)

sc = StandardScaler()

X_train = sc.fit_transform(X_train)

X_test = sc.transform(X_test)

gnb2 = GaussianNB()

nbc = gnb2.fit(X_train, Y_train)

nbc_prediction = nbc.predict(X_test)

f, ax = plt.subplots(figsize=(5,3))

sns.heatmap(confusion_matrix(Y_test, nbc_prediction), annot=True, fmt=".0f",
ax=ax)

plt.show()

print('Akurasi = ', accuracy_score(Y_test, nbc_prediction))

print(classification_report(Y_test, nbc_prediction))
```