

SKRIPSI

**ANALISIS DATA UNTUK MENGETAHUI HUBUNGAN
ANTARA IPK DAN HASIL TPA+**



Muhamad Rifqi Mulyadi

NPM: 2015730038

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2022**

UNDERGRADUATE THESIS

**DATA ANALYSIS TO DETERMINE THE RELATIONSHIP
BETWEEN GPA AND TPA + RESULTS**



Muhamad Rifqi Mulyadi

NPM: 2015730038

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2022**

LEMBAR PENGESAHAN

**ANALISIS DATA UNTUK MENGETAHUI HUBUNGAN
ANTARA IPK DAN HASIL TPA+**

Muhamad Rifqi Mulyadi

NPM: 2015730038

Bandung, 22 Juni 2022

Menyetujui,

Pembimbing

**Digitally signed
by Mariskha Tri
Adithia**

Mariskha Tri Adithia, P.D.Eng

Ketua Tim Penguji

**Digitally signed
by Natalia**

Natalia, M.Si.

Anggota Tim Penguji

**Digitally signed
by Cecilia Esti
Nugraheni**

Dr.rer.nat. Cecilia Esti Nugraheni

Mengetahui,

Ketua Program Studi

**Digitally signed
by Mariskha Tri
Adithia**

Mariskha Tri Adithia, P.D.Eng

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

ANALISIS DATA UNTUK MENGETAHUI HUBUNGAN ANTARA IPK DAN HASIL TPA+

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 22 Juni 2022



Muhamad Rifqi Mulyadi
NPM: 2015730038

ABSTRAK

Prestasi akademik merupakan sebuah bukti keberhasilan proses studi seseorang. Keberhasilan studi mahasiswa pada tahap sarjana dapat diukur dari indeks prestasi kumulatif atau IPK. IPK juga bisa menjadi indikator yang digunakan untuk menilai mutu dari suatu program studi. Maka dari itu analisis terhadap variabel yang dapat mempengaruhi IPK merupakan hal yang penting untuk dilakukan. Ada banyak variabel yang dapat mempengaruhi keberhasilan studi mahasiswa, misalnya kemampuan akademik dan minat terhadap prodi yang dituju. Pada skripsi ini akan dianalisis hubungan antara nilai IPK dan nilai TPA+.

TPA+ adalah salah satu bentuk tes psikologi yang digunakan untuk mengukur kecerdasan intelektual seseorang. TPA+ banyak digunakan dalam berbagai bidang, misalnya proses rekrutmen kerja dan proses seleksi masuk perguruan tinggi. Dalam kasus ini FTIS dan LPH UNPAR mengadakan kegiatan TPA+ untuk, mengetahui potensi akademik mahasiswa. Hasil pengukuran dari TPA+ merupakan bilangan atau skor.

Hubungan antara nilai TPA+ dan nilai IPK dapat diketahui dengan melakukan klasifikasi, lalu di uji keakuratannya. Data akan dikategorikan menjadi 3 kategori berdasarkan IPK, kategori 1 jika IPK mahasiswa kurang dari 2, Kategori 2 jika IPK lebih dari sama dengan 2 dan kurang dari 3, kategori 3 jika IPK mahasiswa lebih dari 3. *Classification* atau klasifikasi adalah sebuah proses untuk mengkategorikan sebuah kumpulan data ke kelasnya atau biasanya disebut target, label atau categories. *classification* memiliki dua langkah. Pertama adalah learning step dimana algoritma akan membuat sebuah model klasifikasi dengan menganalisis set pelatihan atau training set yang terdiri dari kumpulan data dan label kelas yang terkait. Kemudian *classification step* di mana model klasifikasi digunakan untuk memprediksi label kelas untuk data yang baru. Klasifikasi dilakukan dengan algoritma *K-nearest neighbor* atau KNN. Algoritma KNN merupakan algoritma klasifikasi yang mengklasifikasikan data berdasarkan mayoritas dari kategori *k*-tetangga terdekat.

Kemudian akan dibuat juga sebuah perangkat lunak untuk memprediksi kategori IPK berdasarkan nilai TPA+. Pengguna akan memasukkan nilai-nilai TPA+ untuk kemudian diprediksi oleh perangkat lunak, atau pengguna bisa memasukkan sebuah file yang berisi beberapa nilai TPA+ untuk kemudian diprediksi oleh perangkat lunak. Hasilnya disimpan dalam file .xlsx

Berdasarkan pengujian, terdapat keterikatan antara nilai TPA+ dan IPK, namun keterikatan tersebut bukan ikatan yang kuat karena akurasi yang didapat tidak melebihi 0.7. Perangkat lunak juga bisa melakukan prediksi dengan keakuratan 0.66

Kata-kata kunci: Klasifikasi, prediksi *k-nearest neighbor*, akurasi

ABSTRACT

Academic achievement is a proof of the success of one's study process. The success of student studies at the first level can be measured by the achievement index or GPA. GPA can also be an indicator used to assess the quality of a study program. Therefore, an analysis of the variables that can affect GPA is an important thing to do. There are many variables that can affect the success of a student's study, for example academic ability and interest in the intended study program. In this thesis will be analyzed the relationship between the value of GPA and the value of TPA+.

TPA+ is a form of psychological test used to measure a person's intelligence. TPA+ is widely used in various fields, such as job recruitment process and college entrance selection process. In this case, FTIS and LPH UNPAR held a TPA+ activity to find out the academic potential of students. The measurement results from TPA+ are numbers or scores.

The relationship between the TPA+ value and the GPA value can be known by classifying, then testing its accuracy. The data will be categorized into 3 categories based on the GPA, category 1 if the student's GPA is less than 2, Category 2 if the GPA is more than equal to 2 and less than 3, category 3 if the student's GPA is more than 3. *Classification* is a process to categorizes a data set into its classes or usually called targets, labels or categories. classification has two steps. The first is the learning step where the algorithm will create a classification model by analyzing the training set or training set consisting of a data set and associated class labels. Then *classification step* where the classification model is used to predict the class label for the new data. Classification is done by *K-nearest neighbor* or KNN algorithm. The KNN algorithm is a classification that classifies data based on the category of *k*-nearest neighbors.

Then a software will also be made to predict the GPA category based on the TPA+ value. The user will enter TPA+ values for later prediction by the software, or the user can enter a file containing several TPA+ values for later prediction by the software. The result is saved in .xlsx file

Based on this, there is an association between the TPA+ and GPA scores, but the attachment is not strong because the accuracy obtained does not exceed 0.7. The software can also make predictions with an accuracy of 0.66

Keywords: classification, prediction, k-nearest neighbor, accuracy

KATA PENGANTAR

Puji dan syukur penulis ucapkan pada hadirat Allah SWT, karena dengan rahmat dan izin dari-Nya penulis dapat menyelesaikan skripsi ini, yang merupakan salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Teknik Informatika Fakultas Teknologi Informasi dan Sains Universitas Katolik Parahyangan. Penyusunan skripsi ini tidak terlepas dari bantuan berbagai pihak, baik langsung maupun tidak langsung. Secara khusus, penulis ingin berterima kasih kepada:

1. Allah SWT atas segala Rahmat dan Izin-Nya.
2. Keluarga yang selalu memberikan dukungan.
3. Ibu Mariskha Tri Adithia, S.Si., M.Sc., PDEng selaku dosen pembimbing yang membimbing dan membantu dalam proses penulisan skripsi.
4. Ibu Natalia, S.Si, M.Si dan Ibu Dr.rer.nat. Cecilia Esti Nugraheni, ST, MT. selaku dosen penguji yang telah memberikan masukan kritik serta saran terhadap buku skripsi ini, sehingga penulisan buku skripsi ini menjadi lebih baik lagi.
5. Teman-teman dekat dari teknik informatika, dungcer, aslab FE serta seluruh kerabat dekat yang tidak mungkin penulis tulis disini satu per satu untuk dukungan moral yang menyemangati penulis dalam menyelesaikan skripsi ini.

Semoga Allah SWT membalas kebaikan kepada seluruh pihak yang telah memberikan bantuan serta dukungan kepada penulis dalam menyelesaikan skripsi ini. Penulis menyadari bahwa penelitian ini masih jauh dari kata sempurna. Oleh karena itu, penulis memohon maaf jika terdapat kesalahan. Semoga buku skripsi ini dapat memberi informasi yang bermanfaat

Bandung, Juni 2022

Penulis

DAFTAR ISI

| | |
|--|-------------|
| KATA PENGANTAR | xv |
| DAFTAR ISI | xvii |
| DAFTAR GAMBAR | xix |
| DAFTAR TABEL | xxi |
| 1 PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah | 2 |
| 1.3 Tujuan | 2 |
| 1.4 Batasan Masalah | 2 |
| 1.5 Metodologi | 2 |
| 1.6 Sistematika Pembahasan | 2 |
| 2 LANDASAN TEORI | 5 |
| 2.1 Klasifikasi | 5 |
| 2.1.1 Definisi Klasifikasi | 5 |
| 2.1.2 Algoritma KNN | 5 |
| 2.1.3 Contoh Perhitungan Algoritma KNN | 5 |
| 2.2 <i>Library</i> Pendukung | 7 |
| 2.2.1 <i>pandas</i> | 7 |
| 2.2.2 <i>Numpy</i> | 7 |
| 2.2.3 <i>SMOTE</i> | 7 |
| 2.3 TPA+ | 7 |
| 2.4 Evaluasi Klasifikasi | 8 |
| 3 EKSPLORASI TEKNOLOGI | 9 |
| 3.1 <i>Tools</i> dan <i>Library</i> | 9 |
| 3.1.1 <i>Pandas</i> | 9 |
| 3.1.2 <i>Numpy</i> | 10 |
| 3.1.3 <i>Matplotlib</i> | 11 |
| 4 ANALISIS DATA DAN EKSPERIMEN | 13 |
| 4.1 Analisa Masalah | 13 |
| 4.2 Analisa Penyelesaian Masalah | 13 |
| 4.3 Data TPA+ | 13 |
| 4.4 Data Cleaning | 14 |
| 4.5 Normalisasi | 14 |
| 4.6 Klasifikasi Data | 17 |
| 4.6.1 Klasifikasi Data dengan <i>SMOTE</i> | 18 |
| 4.6.2 Klasifikasi Data dengan Atribut Tertentu | 19 |

| | | |
|----------|--|-----------|
| 4.7 | Analisa Perangkat Lunak | 20 |
| 4.7.1 | Analisa Kebutuhan Masukan dan Keluaran | 20 |
| 4.7.2 | Langkah-Langkah Penyelesaian | 21 |
| 4.7.3 | Flowchart Proses Perangkat Lunak | 21 |
| 4.7.4 | Rancangan Kelas | 23 |
| 4.8 | Kebutuhan Masukan dan Keluaran | 23 |
| 4.9 | Perancangan Antarmuka | 24 |
| 4.10 | Diagram kelas | 26 |
| 5 | IMPLEMENTASI DAN PENGUJIAN | 29 |
| 5.1 | Impelementasi Antarmuka | 29 |
| 5.2 | Pengujian Fungsional | 30 |
| 5.2.1 | Pengujian dengan library sklearn | 30 |
| 5.2.2 | Pengujian dengan data iris | 31 |
| 5.2.3 | Pengujian prediksi satu data | 33 |
| 5.2.4 | Pengujian prediksi beberapa data | 33 |
| 6 | PENUTUP | 35 |
| 6.1 | Kesimpulan | 35 |
| 6.2 | Saran | 35 |
| | DAFTAR REFERENSI | 37 |
| | A KODE PROGRAM | 39 |

DAFTAR GAMBAR

| | | |
|------|--|----|
| 3.1 | Gambar contoh plot pada matplotlib | 11 |
| 4.1 | Gambar frekuensi kategori | 16 |
| 4.2 | hasil klasifikasi | 19 |
| 4.3 | frekuensi kelas setelah smote | 20 |
| 4.4 | hasil klasifikasi smote | 21 |
| 4.5 | hasil klasifikasi TKD 5 dan 6 | 22 |
| 4.6 | Gambar flowchart | 22 |
| 4.7 | Gambar rancangan kelas | 23 |
| 4.8 | contoh input | 24 |
| 4.9 | contoh input | 24 |
| 4.10 | rancangan GUI | 25 |
| 4.11 | diagram kelas | 26 |
| 5.1 | Impelementasi Antarmuka | 29 |
| 5.2 | dialog pilihan file | 30 |
| 5.3 | hasil test | 33 |
| 5.4 | file input test.csv | 33 |
| 5.5 | pilih file | 34 |
| 5.6 | hasil test file | 34 |

DAFTAR TABEL

| | | |
|-----|--------------------|---|
| 2.1 | Tabel contoh knn 1 | 6 |
| 2.2 | Tabel contoh knn 2 | 6 |
| 2.3 | Tabel contoh knn 2 | 6 |
| 2.4 | Tabel contoh knn 3 | 6 |

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Prestasi akademik merupakan sebuah bukti keberhasilan proses studi seseorang. Keberhasilan studi mahasiswa pada tahap sarjana dapat diukur dari indeks prestasi kumulatif atau IPK. IPK juga bisa menjadi indikator yang digunakan untuk menilai mutu dari suatu program studi. Maka dari itu analisis terhadap variabel yang dapat mempengaruhi IPK merupakan hal yang penting untuk dilakukan. Ada banyak variabel yang dapat mempengaruhi keberhasilan studi mahasiswa, misalnya kemampuan akademik dan minat terhadap prodi yang dituju.

Setiap tahun, FTIS dan LPH UNPAR mengadakan kegiatan TPA+ bagi mahasiswa baru. TPA+ adalah salah satu bentuk tes psikologi yang digunakan untuk mengukur kecerdasan intelektual seseorang. TPA+ banyak digunakan dalam berbagai bidang, misalnya proses rekrutmen kerja dan proses seleksi masuk perguruan tinggi. Dalam kasus ini FTIS dan LPH UNPAR mengadakan kegiatan TPA+ untuk, mengetahui potensi akademik mahasiswa. Hasil pengukuran dari TPA+ merupakan bilangan atau skor.

Data TPA+ terdiri dari 5 nilai tes, yaitu nilai tes TKD1, TKD2, TKD5, TKD6 dan nilai tes logika verbal. Tes TKD 1 (Comprehension) adalah sebuah rangkaian tes mengenai pengertian umum yang terdiri dari 30 soal dengan pilihan jawaban a, b, c atau d. Tes ini bertujuan untuk menilai kecerdasan sosial yaitu sikap terhadap situasi sosial praktis. Tes TKD 2 (Information) adalah sebuah rangkaian tes menyelesaikan kalimat yang terdiri dari 40 soal. Mahasiswa diminta untuk memilih satu kata dari 4 kata untuk menyelesaikan suatu kalimat. Tes ini bertujuan untuk menilai ruang lingkup pengetahuan, kewaspadaan (allertness) seseorang terhadap dunia luar. Tes TKD 5 (Arithmetic) terdiri dari 20 soal hitungan. TKD 6 (Deret Angka) adalah sebuah tes yang terdiri dari 25 soal deret angka dan mahasiswa diminta untuk melanjutkan 2 bilangan yang sesuai dengan irama deret. Tes ini bertujuan untuk mengukur penalaran berhitung dengan angka dan mengukur logika berpikir. Terakhir adalah tes logika verbal adalah sebuah tes yang terdiri dari 15 soal berupa soal cerita dan memilih jawaban a, b, c atau d Tes ini bertujuan untuk menilai cara berpikir logis (hal ini ada kaitannya dengan tugas kuliah; bagaimana menyusun kalimat dan membahasakannya dalam struktur s-p-o-k).

Data IPK terdiri dari data nilai IPK mahasiswa, data yang digunakan merupakan data IPK mahasiswa tahun 2017 dan 2018 jurusan teknik informatika, matematika dan fisika UNPAR. Pada skripsi ini, akan dianalisis hubungan antara hasil TPA+ dengan IPK. Sebelum dilakukan analisis Data IPK dan data hasil TPA+ digabungkan dan dikategorikan sesuai dengan IPK. Kategori 1 rentang IPK kurang dari 2, kategori 2 rentang IPK lebih dari sama dengan 2 sampai kurang dari 3, kategori 3 rentang IPK lebih dari 3.

Classification atau klasifikasi adalah sebuah proses untuk mengkategorikan sebuah kumpulan data ke kelasnya atau biasanya disebut target, label atau categories. classification memiliki dua langkah. Pertama adalah learning step dimana algoritma akan membuat sebuah model klasifikasi dengan menganalisis set pelatihan atau training set yang terdiri dari kumpulan data dan label kelas yang terkait. Kemudian classification step di mana model klasifikasi digunakan untuk memprediksi label kelas untuk data yang baru. Algoritma klasifikasi yang digunakan adalah algoritma *k-nearest*

neighbor (KNN). Algoritma k-nearest neighbor adalah algoritma klasifikasi yang bertujuan untuk mengklasifikasikan dan memprediksi data berdasarkan mayoritas dari data tetangga terdekat

Penerapan teknik *classification* bertujuan untuk mengkategorikan data, serta pembuatan model klasifikasi yang bisa digunakan untuk memprediksi kategori dari data yang baru. Pada skripsi ini akan dibuat perangkat lunak untuk mengklasifikasi dan memprediksi kategori IPK mahasiswa berdasarkan hasil tes TPA+. Pengguna akan memasukkan nilai hasil tes TPA+ untuk kemudian diprediksi oleh perangkat lunak, atau pengguna bisa memasukkan sebuah file yang berisi beberapa nilai TPA+ untuk kemudian diprediksi oleh perangkat lunak. Hasilnya disimpan dalam file .xlsx

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, ditemukan rumusan masalah sebagai berikut:

1. Bagaimana cara mengidentifikasi hubungan antara nilai TPA+ dan IPK?
2. Bagaimana cara memprediksi data TPA+ baru?

1.3 Tujuan

Penelitian ini dilakukan dengan tujuan sebagai berikut:

1. Membangun model klasifikasi terkait dan menerapkannya ke data TPA+ dan IPK.
2. Membangun perangkat lunak yang bisa digunakan untuk mengklasifikasi data TPA+ dan IPK baru.

1.4 Batasan Masalah

Nilai yang digunakan hanya nilai IPK dan TPA+ mahasiswa jurusan fisika, matematika dan teknik informatika.

1.5 Metodologi

Metodologi yang dilakukan dalam melakukan penelitian ini adalah sebagai berikut:

1. Melakukan studi literatur
Studi literatur dilakukan untuk mempelajari beberapa teori yang dibutuhkan dalam melakukan penelitian ini. Teori-teori yang dibutuhkan adalah teori mengenai klasifikasi, algoritma *knn*, *python* beserta *library* pendukung.
2. Membangun perangkat lunak
Proses pembangunan perangkat lunak dimulai dengan membaca data dan melakukan klasifikasi. Kemudian perangkat lunak membaca input dari pengguna berupa nilai hasil tes TPA+ dan kemudian perangkat lunak memprediksi kategori dari IPK.
3. Melakukan pengujian terhadap perangkat lunak
Pengujian perangkat lunak dilakukan untuk menguji kebenaran hasil analisis dan kinerja dari perangkat lunak.
4. Menulis dokumen
Penulisan dokumen akan dibagi menjadi 6 bab utama yang akan dijelaskan pada subbab Sistematika Pembahasan.

1.6 Sistematika Pembahasan

Pada laporan penelitian ini akan terdapat enam bab yang ditulis secara sistematis. Enam bab tersebut dapat dijelaskan secara singkat sebagai berikut:

-
1. Bab 1 merupakan bab pendahuluan yang berisi latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi penelitian, dan sistematika pembahasan.
 2. Bab 2 berisi dasar teori. Bab ini berisi dasar teori tentang klasifikasi, algoritma knn, *library python* yang akan digunakan, data TPA+, serta evaluasi klasifikasi.
 3. Bab 3 berisi eksplorasi teknologi. Bab ini berisi penjelasan *tools dan library* yang akan digunakan.
 4. Bab 4 berisi Analisis data dan Eksperimen. Bab ini berisi analisa data yang dilakukan, serta perancangan perangkat lunak.
 5. Bab 5 merupakan bab implementasi dan pengujian perangkat lunak. Pada bab ini akan dicantumkan potongan kode program, penjelasan cara melakukan pengujian perangkat lunak, dan kesimpulan hasil pengujian.
 6. Bab 6 berisi kesimpulan dari hasil penelitian ini dan saran yang mungkin akan berguna untuk penelitian-penelitian selanjutnya