SKRIPSI

ANALISIS DATA HASIL TES PISA MENGGUNAKAN PENAMBANGAN DATA

Gavin Weldi Kusmana

NPM: 6181901001

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2023

UNDERGRADUATE THESIS

UTILIZATION OF DATA MINING FOR ANALYZING PISA TEST RESULT

Gavin Weldi Kusmana

NPM: 6181901001

DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2023

LEMBAR PENGESAHAN

ANALISIS DATA HASIL TES PISA MENGGUNAKAN PENAMBANGAN DATA

Gavin Weldi Kusmana

NPM: 6181901001

Bandung, 5 Juli 2023

Menyetujui,

Pembimbing

Digitally signed by Mariskha Tri Adithia

Mariskha Tri Adithia, P.D.Eng

Ketua Tim Penguji

Digitally signed by Vania Natali

Vania Natali, M.T.

Anggota Tim Penguji

Digitally signed by Rosa de Lima E. Padmowati

Rosa De Lima, M.T.

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 HASIL TES PISA MENGGUNAKAN PENAMBANGAN DATA

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 5 Juli 2023

Gavin Weldi Kusmana
NPM: 6181901001

ABSTRAK

PISA (*Programme for International Student Assessment*) merupakan program yang dilaksanakan tiga tahun sekali dengan tujuan untuk menguji kemampuan siswa dari berbagai negara termasuk Indonesia. Pada tahun 2015, negara Indonesia menempati peringkat 64 dari 70 negara yang berpartisipasi pada tes PISA. Dengan begitu menunjukkan adanya kebutuhan untuk menganalisis faktor-faktor yang berhubungan dengan nilai tes seorang siswa. Diharapkan dengan tercapainya penelitian ini yaitu mendapatkan faktor-faktor yang berhubungan dengan nilai tes PISA, maka nilai pada tes PISA selanjutnya dapat meningkat. Untuk mengetahui faktor-faktor yang berhubungan nilai tes siswa, pada penelitian ini menggunakan teknik penambangan data.

Metode yang digunakan adalah dengan pendekatan klasterisasi untuk mencari pola-pola dari setiap klaster yang dapat memberikan wawasan tentang faktor-faktor yang berhubungan hasil tes PISA. Sebelum menganalisis, diperlukan data tes PISA dengan cara mengunduh melalui situs resmi PISA. Kemudian dilakukan terlebih dahulu eksplorasi dan analisis data guna mencari tahu sifat dari data serta diharapkan untuk mendapat kesimpulan sementara untuk mempersempit arah analisis. Dilanjutkan dengan teknik pada statistika inferensi dalam bentuk uji inferensi untuk memilih fitur yang memiliki korelasi dengan nilai serta membuang fitur yang tidak memiliki korelasi dengan nilai. Proses tersebut menggunakan metode *chi-square* dan ANOVA dengan alasan tipe data dari hasil tes PISA selain tes kognitif merupakan kategorik yaitu berupa kuesioner. Selanjutnya, melakukan analisis data menggunakan dua jenis algoritma pengelompokan untuk mengidentifikasi pola berdasarkan nilai rata-rata tes PISA. Algoritma pengelompokan yang dipakai adalah *K-Means* dan *Agglomerative* dengan alasan ingin membandingkan hasil dari dua cara kerja yang berbeda. Dilakukan pengelompokan berdasarkan negara dan siswa untuk memperoleh wawasan dari kedua perspektif.

Hasil penelitian ini mengungkapkan empat faktor yang signifikan dalam hubungannya dengan nilai siswa. Faktor-faktor tersebut adalah kualitas material laboratorium yang baik, penyediaan ruang belajar yang memadai untuk siswa mengerjakan pekerjaan rumah (PR), tersedianya staf pengajar yang dapat membantu siswa dalam mengerjakan PR, serta penggunaan tes buatan guru untuk meningkatkan kurikulum. Hasil dari penelitian ini merupakan informasi berharga yang dalam penyampaiannya menggunakan *dashboard* sebagai media. Dalam kesimpulannya, penelitian ini memberikan bukti bahwa faktor-faktor tersebut berpotensi berhubungan secara signifikan terhadap hasil tes PISA seorang siswa. Informasi ini dapat digunakan sebagai dasar untuk mengembangkan strategi dan kebijakan pendidikan yang lebih efektif guna meningkatkan kualitas pendidikan.

Kata-kata kunci: analisis hasil tes PISA, teknik pengelompokan, penambangan data

ABSTRACT

PISA (Program for International Student Assessment) is a program that held every three years with the aim of testing the ability of students from various countries including Indonesia. In 2015, the country of Indonesia occupies the position ranked 64th out of 70 countries participating in the PISA test. Therefore indicates the need to analyze the related factors with a student's test score. It is hoped that with the achievement of this research, namely get the factors related to the PISA test score, then the score on the next PISA test can increase. To find out which factors related to student test scores, in this study using data mining techniques.

The method used is a clustering approach to look for patterns patterns of each cluster which can provide insight into which factors results related to the PISA test. Before analyzing, it is necessary to use PISA test data how to download through the official PISA website. Then do it first exploration and analysis of data to find out the nature of the data and what is expected of it draw temporary conclusions to narrow the direction of analysis. Next with techniques on statistical inference in the form of inference tests to choose features that have a correlation with values and highlight features that do not has a correlation with PISA test score. The process uses the chi-square method and ANOVA on the grounds that the data type of PISA is categorical, namely in the form of questionnaire. Next, perform data analysis using two types of clustering algorithms to identify patterns based on the average score of the PISA test. The clustering algorithm used is K-Means and Agglomerative with the reason for wanting to compare the results of the two different ways of working. Done grouping by country and students to gain insight from both perspectives.

The results of this study reveal four significant factors in relation to relation to student grades. These factors are the quality of the material a good laboratory, providing adequate study space for students to do homework, the availability of provider staff who can help students in doing homework, as well as the use of teacher-made tests to improve school curriculum. The results of this study are valuable information in the withdrawal using the dashboard as a medium. In confusion, this study provides evidence that these factors have potential significantly related to the results of a student's PISA test. Information This can be used as a basis for developing strategies and policies more effective education to improve the quality of education.

Keywords: analysis of PISA test results, clustering techniques, data mining

KATA PENGANTAR

Puji dan syukur peneliti panjatkan kepada Tuhan Yang Maha Esa, atas segala rahmat yang dilimpahkan oleh-Nya, sehingga penyusunan Skripsi dapat diselesaikan. Skripsi ini merupakan sebuah penelitian dengan judul Analisis Data Hasil Tes PISA Menggunakan Penambangan Data. Skripsi ini diajukan sebagai syarat untuk menempuh pendidikan di Program Studi Teknik Informatika, Fakultas Teknologi Informasi dan Sains, Universitas Katolik Parahyangan, Bandung.

Dalam penyusunan dokumen Skripsi ini tak lupa peneliti memberikan ucapan terima kasih kepada pihak-pihak yang terlibat baik dalam memberi bimbingan, bantuan dan dukungan secara langsung atau tidak ke dalam proses penyusunan penelitian, terutama untuk:

1. Natalia, M.Si., selaku dosen pembimbing penelitian yang berjudul Analisis Data Hasil Tes PISA Menggunakan Penambangan Data.
2. Vania Natali, M.T., dan Rosa De Lima, M.T., selaku dosen-dosen penguji.
3. Orangtua, yang selalu mendukung dan menemani peneliti dalam penyusunan dokumen Skripsi ini.
4. Stefanny Abigail serta Gabut Club yang selalu mendukung dan menemani penulis dalam penyusunan dokumen Skripsi ini.

Akhir kata peneliti berharap agar Skripsi ini dapat memberikan sumbangan nyata ataupun pembelajaran bagi kemajuan Teknik Informatika dan bagi pihak yang memerlukannya.

Bandung, Juni 2023

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR KODE PROGRAM

BAB 1

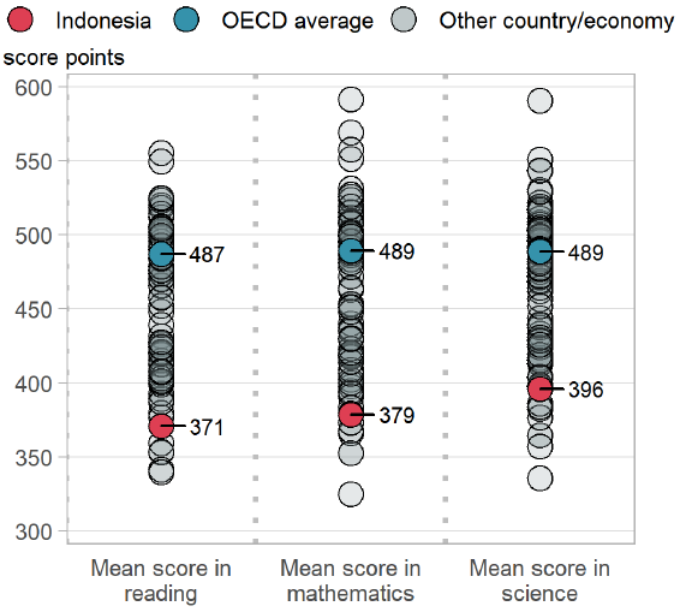
PENDAHULUAN

1.1 Latar Belakang

PISA (*Programme for International Student Assessment*) adalah sebuah program yang diselenggarakan oleh organisasi internasional OECD (*Organisation for Economic Co-operation and Development*) untuk mengevaluasi kemampuan siswa di berbagai negara. Program ini mengukur kemampuan siswa dalam tiga area literasi, yaitu keterampilan membaca, matematika, dan sains. Tujuan dari literasi membaca adalah untuk mengetahui keahlian siswa dalam memahami, menanggapi, dan merefleksikan dalam konteks membaca teks. Tujuan literasi matematika adalah untuk mengetahui kemampuan bernalar siswa secara matematis dalam menggunakan konsep, prosedur, fakta, dan perangkat matematis ketika mendeskripsikan, menjelaskan, serta memprediksi fenomena. Tujuan dari literasi sains adalah untuk mengetahui kemampuan siswa dalam menanggapi isu-isu sains dengan menggunakan gagasan-gagasan ilmiah.

Tes PISA dilakukan setiap tiga tahun sekali sejak tahun 2000 dengan subjek yaitu siswa yang berusia 15 tahun, alasannya adalah pada usia ini kaum muda di sebagian besar negara mendekati akhir wajib belajar. Setiap diselenggarakannya tes akan ada satu literasi yang menjadi fokus pengetesan, dalam hal ini berarti perbandingan nilai pada literasi tertentu dapat diketahui dalam sembilan tahun. Sebagai contoh, literasi yang pertama kali diselenggarakan pada tahun 2000 dengan fokus literasi membaca kemudian pada tahun 2003 fokus bidang matematika dan pada tahun 2006 berfokus pada literasi sains. Tahun 2009 merupakan pengetesan yang memfokuskan pada literasi membaca, maka dari itu memakan sembilan tahun untuk mengetahui perbandingan nilainya. Dua subjek lainya yang tidak menjadi fokus pengetesan dikaji sebagai subjek pendamping. Pemilihan sekolah dan siswa dilakukan dengan mementingkan inklusivitas, sehingga sampel siswa berasal dari berbagai latar belakang dan kemampuan. Beberapa sekolah dan siswa yang terdaftar dapat dikeluarkan dari sampel tes PISA karena PISA menetapkan sebuah pengecualian yaitu tidak boleh mewakili lebih dari lima persen dari populasi target. Dalam batas itu, siswa dapat dikeluarkan karena berbagai alasan, termasuk keterpencilan dan tidak dapat diaksesnya sekolah, cacat intelektual atau fisik, atau kurangnya kemahiran dalam bahasa ujian. Hingga tahun 2022, tes PISA telah melibatkan lebih dari 90 negara dan 3.000.000 siswa.

Negara Indonesia telah ikut serta dalam tes PISA sejak tahun 2000. Setelah 15 tahun Indonesia menjadi partisipan dalam tes PISA, tetapi belum pernah mengalami peningkatan peringkat yang signifikan. Berikut merupakan hasil tes PISA Indonesia dari tiga tahun ke tiga tahun, tahun 2000 menempati peringkat 39 dari 41 negara, tahun 2003 menempati peringkat 55 dari 57 negara, tahun 2006 menempati peringkat 54 dari 57 negara, tahun 2009 menempati peringkat 69 dari 74 negara, tahun 2012 menempati peringkat 64 dari 65 negara, dan hasil tes PISA tahun 2015 mendapati peringkat 62 dari 70 negara. Kesimpulan yang didapat adalah peringkat yang didapat sering kali 10 terbawah dunia. Pada Gambar 1.1, nilai PISA Indonesia melalui data terakhir yaitu tahun 2018 lebih rendah daripada nilai rata-rata negara *OECD*. Berdasarkan artikel Pratiwi, I. (2019) [1], hasil asesmen PISA di Indonesia dapat dikategorikan buruk, maka pemerintah Indonesia selalu mendapat tekanan dari publik karena dianggap belum berhasil dalam membuat sistem pembelajaran yang baik.



Gambar 1.1: Perbandingan kinerja negara Indonesia dalam membaca, matematika dan sains tahun 2018[2].

Selain menguji kemampuan siswa, panitia OECD memberikan kuesioner kepada sekolah dan siswa memuat pertanyaan tentang hal-hal yang berhubungan dengan latar belakang dan kondisi dari siswa serta sekolah yang mengikuti tes PISA. Informasi-informasi dari kuesioner tersebut memberikan tiga tipe hasil utama penilaian tes PISA sebagai berikut.

1. Tipe pertama yaitu beberapa indikator seperti tingkat kepuasan hidup siswa, emosi positif dan negatif, keinginan untuk melanjutkan pendidikan ke jenjang yang lebih tinggi seperti sarjana dan magister.
2. Tipe kedua indikator-indikator yang terbentuk dari hasil kuesioner yang menunjukkan bagaimana keterampilan yang dimiliki siswa berkaitan dengan berbagai variabel demografi, sosial ekonomi, dan pendidikan, serta hasil dari pendidikan yang lebih luas seperti pencapaian tingkat pendidikan dan kesejahteraan.
3. Tipe ketiga yaitu indikator-indikator dalam tren, dimulai dengan keikutsertaan negara untuk ketujuh kalinya dalam tes PISA.

Terfokus pada tipe ketiga yaitu dapat menunjukkan perubahan nilai berdasarkan *median* hasil tes PISA. Pemakaian nilai tengah karena agar nilai tidak terpengaruh dengan nilai ekstrem. Tidak hanya nilai tes PISA, terdapat juga variasi hasil di antara siswa, dalam hubungannya antara hasil dengan berbagai variabel khusus siswa, sekolah, dan sistem. Seperti yang sudah disebutkan diawal bahwa PISA tersedia hanya sampai 2018, penelitian ini akan berfokus pada PISA tahun 2015. Alasannya adalah karena tahun 2015 merupakan pengetesan PISA dengan fokus literasinya adalah sains. Menurut buku *The Case for STEM Education* bab 4 [2], literasi sains merupakan kemampuan siswa untuk memahami dan menggunakan pengetahuan, konsep, dan keterampilan sains dalam pemecahan masalah, pengambilan keputusan, dan partisipasi dalam kehidupan sehari-hari. Dengan penjelasan tersebut diasumsikan bahwa dengan menganalisis nilai literasi sains dapat mewakili tingkat akademik dari suatu negara karena pengaruh dari literasi sains yang paling dapat merefleksikan kehidupan akademik.

Dari survei yang dikeluarkan oleh panitia OECD dapat ditemukan hal yang menarik dari indikator-indikator yang tersedia. Proses penambangan data merupakan salah satu teknik yang dapat digunakan untuk menemukan pola tertentu yang menarik dan juga *insights* dari pola-pola yang

[2]OECD, *Programme For International Student Assessment (PISA) Result From PISA 2018*, diakses pada 27 Januari 2023, https://www.oecd.org/pisa/publications/PISA2018_CN_IDN.pdf

didapat. Pola-pola yang mungkin akan dihasilkan dari proses analisis adalah seperti karakterisasi dan diskriminasi, klasifikasi dan regresi dengan tujuan memprediksi, analisis klaster, dan lain-lain. Untuk mendapatkan pola tersebut dilakukan dengan menggunakan banyak metode seperti contohnya statistik yang mempelajari kumpulan, analisis, interpretasi atau penjelasan, penyajian dari data. Teknik *clustering* juga dapat dipilih menjadi metode analisis guna mencari pola atau faktor yang berhubungan dengan tes PISA. *Machine learning* dapat digunakan untuk membantu mengenali pola yang kompleks dan membuat keputusan cerdas berdasarkan data. Kinerja dari *machine learning* tersebut diharapkan akan menghasilkan pola yang unik, misal dapat menentukan bagaimana status ekonomi/GDP dari negara mempengaruhi nilai tes PISA. Adapun hasil lain yaitu dapat mengelompokkan negara mana saja yang nilai tes PISA tinggi berdasarkan latar belakang dari siswa. Tahapan terakhir dari penelitian ini adalah pembuatan perangkat lunak yang dapat mengaplikasikan model *machine learning* berdasarkan eksperimen.

1.2 Rumusan Masalah

Rumusan masalah yang muncul berdasarkan deskripsi dan latar belakang yang sudah dibahas adalah sebagai berikut:

1. Bagaimana sistem pelaksanaan tes PISA?
2. Bagaimana cara untuk menyiapkan *dataset* tes PISA agar siap untuk dianalisis?
3. Bagaimana tahapan *data mining* dilakukan untuk mengekstrak *insights* dari data tes PISA?
4. Bagaimana cara menampilkan *insights* yang merupakan hasil analisis dari penelitian ini?

1.3 Tujuan

1. Melakukan pembelajaran mengenai tes PISA
2. Melakukan eksplorasi terkait cara penyiapan data tes PISA
3. Melakukan proses *data mining* untuk mendapatkan informasi berharga dari *dataset* tes PISA
4. Membuat *dashboard* sebagai media penyampaian hasil analisis

1.4 Batasan Masalah

Batasan masalah yang membatasi penelitian adalah sebagai berikut :

1. *Dataset* yang digunakan merupakan data hasil tes PISA tahun 2015
2. Analisis dilakukan terhadap fitur sekolah
3. Algoritma *clustering* yang digunakan adalah *K-Means* dan *Agglomerative*

1.5 Metodologi

Metodologi yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Melakukan studi literatur mengenai *domain knowladge* dari tes PISA
2. Melakukan eksplorasi dan studi literatur tentang statistika deskriptif dan inferensi
3. Melakukan eksplorasi dan studi literatur tentang teknik visualisasi yang ingin dipakai
4. Melakukan eksplorasi dan studi literatur tentang *clustering*
5. Melakukan eksplorasi dan studi literatur tentang klasifikasi
6. Melakukan eksplorasi dan proses pengumpulan data terhadap *dataset* tahun 2015
7. Membuat perangkat lunak yang dapat melakukan penyiapan data, analisis data, dan menampilkan hasil dari analisisnya
8. Melakukan analisis terhadap hasil dari perangkat lunak, merumuskan hasil analisis dan memaparkan hasilnya dengan teknik visualisasi yang sesuai

1.6 Sistematika Pembahasan

Sistematika penulisan pada penelitian ini adalah sebagai berikut:

1. Bab Pendahuluan:
   Pada Bab Pendahuluan akan membahas tentang pengantar penelitian yang berisikan masalah secara umum dan langkah-langkah yang dilakukan dalam menyelesaikan masalah, batasan penelitian, serta tahapan eksperimen yang dilakukan.
2. Bab Landasan Teori:
   Pada Bab Landasan Teori akan dijelaskan tentang bahasan studi yang mencakup keseluruhan penelitian.
3. Bab Analisis Penyelesaian Masalah:
   Pada Bab Analisis Penyelesaian Masalah berisi deskripsi masalah yang akan diselesaikan dan eksperimen menggunakan data yang berukuran kecil terhadap metode-metode yang dipakai pada penelitian ini.
4. Bab Penambangan Data:
   Bab Penambangan Data berisi seluruh tahapan *data science* sebagai sarana penyampaian proses penyelesaian masalah.
5. Bab Peluncuran Model dan Pengujian:
   Bab Peluncuran Model dan Pengujian berisi hasil dari penelitian yang dikaji dalam bentuk *dashboard*.
6. Bab Kesimpulan dan Saran:
   Bab Kesimpulan dan Saran berisi tentang kesimpulan dari seluruh rangkaian dari penelitian ini serta saran yang dapat dilakukan untuk mengembangkan penelitian ini.