

SKRIPSI

EKSPLORASI DAN ANALISIS DATA GEM



Bernadus Lukas

NPM: 6181801077

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



**UNDERGRADUATE THESIS**

**EXPLORATION AND ANALYSIS OF GEM DATASET**



**Bernadus Lukas**

**NPM: 6181801077**

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



**LEMBAR PENGESAHAN**

**EKSPLORASI DAN ANALISIS DATA GEM**

**Bernadus Lukas**

**NPM: 6181801077**

**Bandung, 28 Juni 2022**

**Menyetujui,**

**Pembimbing**

**Digitally signed  
by Cecilia Esti  
Nugraheni**

**Dr.rer.nat. Cecilia Esti Nugraheni**

**Ketua Tim Penguji  
Digitally signed  
by Maria V.  
Claudia M.**

**Maria Veronica, M.T.**

**Anggota Tim Penguji  
Digitally signed  
by Veronica Sri  
Moertini**

**Dr. Veronica Sri Moertini**

**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:

### EKSPLORASI DAN ANALISIS DATA GEM

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 28 Juni 2022



Bernadus Lukas  
NPM: 6181801077





## ABSTRAK

Kegiatan wirausaha memegang peranan penting dalam pembangunan suatu negara atau dapat dikatakan bahwa kewirausahaan salah satu penggerak roda ekonomi suatu negara. Karena orang-orang mulai sadar betapa pentingnya kegiatan wirausaha disuatu negara, maka pada bulan September 1997 *Global Entrepreneurship Monitor* (GEM) didirikan oleh Michael Hay dari London Business School (LBS) dan Bill Bygrave dari Babson College. *Global Entrepreneurship Monitor* (GEM) merupakan sebuah organisasi non-profit yang berfokus pada kewirausahaan. Setiap tahunnya GEM membuat sebuah laporan terkait kewirausahaan di suatu negara. Laporan tersebut dibuat berdasarkan dua buah survei yang diadakan oleh GEM. Survei tersebut adalah *Adult Population Survey* (APS) dan *National Expert Survey* (NES). Data yang dihasilkan dalam survey APS dan NES, dapat dimanfaatkan jika kita dapat mencari informasi berharga di dalam data tersebut. Untuk memperoleh informasi berharga tersebut dapat menggunakan teknik *Data Mining*.

*Data Mining* merupakan sebuah metode untuk menggali informasi dari sekumpulan data yang jumlahnya besar. *Data Mining* memiliki beberapa teknik yaitu regresi, klasifikasi, dan *clustering* yang akan digunakan pada penelitian ini. Sebagai contoh dengan teknik klasifikasi, data GEM untuk tiap negara akan dimanfaatkan untuk memprediksi kelas berdasarkan atribut yang ada pada hasil survei APS dan NES. Kemudian, dengan menggunakan teknik *clustering*, data GEM untuk tiap negara akan dikelompokkan berdasarkan kemiripan atribut pada hasil survei APS dan NES.

Pada penelitian ini dihasilkan perangkat lunak yang mampu mengimplementasikan tahapan *Data Mining* yaitu membersihkan data, melakukan integrasi data, mencari atribut yang berkorelasi, melakukan transformasi data, dan mampu menggunakan teknik-teknik *Data Mining* untuk data GEM. Hasil penelitian ini juga dicantumkan di sebuah situs yang dibuat untuk menampilkan hasil eksplorasi dan analisis data GEM.

Berdasarkan eksplorasi dan analisis data GEM yang dilakukan, disimpulkan bahwa faktor yang paling memiliki pengaruh pada data APS dalam menentukan seberapa tinggi atau rendah nilai *Perceived Capabilities* adalah *Perceived opportunities*, faktor yang paling memiliki pengaruh pada data NES dalam menentukan seberapa tinggi atau rendah nilai *Governmental programs* adalah *Governmental support and policies*.

**Kata-kata kunci:** GEM, APS, NES, *Data Mining*, Regresi, Klasifikasi, *Clustering*



## ABSTRACT

Entrepreneurial activities play an important role in the development of a country or it can be said that entrepreneurship is one of the driving forces of a country's economy. As people began to realize how important entrepreneurial activities are in a country, then in September 1997 Global Entrepreneurship Monitor(GEM) was founded by Michael Hay of the London Business School(LBS) and Bill Bygrave of Babson College. Global Entrepreneurship Monitor(GEM) is a non-profit organization focused on entrepreneurship. Every year GEM makes a report related to entrepreneurship in a country. The report was based on two surveys conducted by GEM. The surveys are Adult Population Survey(APS) and National Expert Survey(NES). The data generated in the APS and NES surveys, can be utilized if we can find valuable information in the data. To obtain such valuable information can use the Data Mining technique.

Data Mining is a method of extracting information from a large set of data. Data Mining has several techniques namely regression, classification, and clustering that will be used in this study. For example, with classification techniques, GEM data for each country will be used to predict classes based on attributes in the APS and NES survey results. Then, using the clustering technique, the GEM data for each country will be grouped based on the similarity of attributes in the APS and NES survey results.

In this study, software was develop that is able to implement the stages of data mining, namely cleaning data, integrating data, looking for correlated attributes, transforming data, and being able to use data mining techniques for GEM data. The results of this study are also listed on a site created to display the results of gem data exploration and analysis.

Based on the exploration and analysis of GEM data carried out, it was concluded that the factor that has the most influence on APS data in determining how high or low the value of Perceived Capabilities is Perceived opportunities, the factor that has the most influence on NES data in determining how high or low the value of Governmental programs is Governmental support and policies.

**Keywords:** GEM, APS, NES, Data Mining, Regression, Classification, Clustering



*Skripsi ini saya persembahkan untuk keluarga tercinta, teman-teman  
dan diri sendiri*



## KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yesus karena atas penyertaanNya penulis dapat menyelesaikan skripsi dengan judul "Eksplorasi dan Analisis Data GEM". Penulisan skripsi ini merupakan salah satu persyaratan dalam mendapatkan gelar sarjana pada Program Studi Informatika, Fakultas Teknologi Informasi dan Sains, Universitas Katolik Parahyangan. Penulis menyadari bahwa skripsi ini tidak mungkin dapat diselesaikan dengan baik tanpa dukungan baik dalam bentuk materi, dan moral. Oleh sebab itu penulis ingin mengucapkan terima kasih kepada:

1. Kedua orang tua, Bapak Ade Chandra Kurniawan dan Ibu Ely, Saudara Brenda Fedora dan Christy Fedora, dan Paman Hendro Susanto yang telah memberikan dukungan materil dan immateril kepada Penulis selama masa kuliah.
2. Ibu Dr.rer.nat. Cecilia Esti Nugraheni, selaku dosen pembimbing yang mendampingi penulis dalam bentuk bimbingan, saran, dan nasihat dalam penulisan skripsi ini.
3. Ibu Maria Veronica, M.T., selaku penguji utama dan Ibu Dr. Veronica Sri Moertini, selaku penguji pendamping yang telah memberikan masukan, kritik, dan saran untuk perbaikan skripsi ini.
4. Seluruh Bapak/Ibu dosen Informatika yang telah memberikan pengetahuan dan pelajaran selama penulis berkuliah di UNPAR.
5. Teman-teman di *server* Discord "server kunci" : Clement, Julyus, Jason, Michael, Ivan, Ks, dan Bill yang telah memberikan hiburan dan dukungan selama proses perkuliahan.
6. Teman-teman proyek Data Science: Kezia, Afi, dan Josie yang telah memberikan hiburan dan dukungan selama proses perkuliahan.
7. Indra Permana Sugianto yang telah membantu penulis dalam memberikan *insight* dalam penulisan skripsi. Gian, Deddy, dan Upekkha yang telah memberikan bantuan selama masa perkuliahan. Serta seluruh keluarga dan teman-teman selama perkuliahan yang telah memberikan semangat dan dukungan yang tidak dapat disebutkan satu per satu.

Akhir kata, semoga Tuhan membalas kebaikan seluruh pihak yang telah membantu penulis dalam menyelesaikan perkuliahan S1 dan penulisan skripsi ini. Dalam penulisan skripsi ini, penulis menyadari bahwa skripsi ini jauh dari sempurna karena keterbatasan pengetahuan penulis sehingga masih memiliki banyak ruang untuk dikembangkan. Oleh karena itu apabila terdapat kesalahan dalam penelitian dan penulisan, penulis meminta maaf dan berharap dapat diberikan kritik dan saran yang membantu penyempurnaan penelitian ini. Semoga skripsi ini dapat bermanfaat bagi semua pembaca.

Bandung, Juni 2022

Penulis





# DAFTAR ISI

<b>KATA PENGANTAR</b>	<b>xv</b>
<b>DAFTAR ISI</b>	<b>xvii</b>
<b>DAFTAR GAMBAR</b>	<b>xix</b>
<b>DAFTAR TABEL</b>	<b>xxi</b>
<b>DAFTAR KODE PROGRAM</b>	<b>xxiii</b>
<b>1 PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang . . . . .	1
1.2 Rumusan Masalah . . . . .	2
1.3 Tujuan . . . . .	2
1.4 Batasan Masalah . . . . .	3
1.5 Metodologi . . . . .	3
1.6 Sistematika Pembahasan . . . . .	3
<b>2 LANDASAN TEORI</b>	<b>5</b>
2.1 GEM . . . . .	5
2.1.1 APS . . . . .	6
2.1.2 NES . . . . .	8
2.2 Data Mining . . . . .	9
2.2.1 Regresi . . . . .	11
2.2.2 Klasifikasi . . . . .	13
2.2.3 Clustering . . . . .	15
2.3 WordPress . . . . .	19
2.4 Python . . . . .	19
2.5 Tableau . . . . .	20
<b>3 STUDI EKSPLORASI ALGORITMA DATA MINING</b>	<b>21</b>
3.1 Analisis Masalah . . . . .	21
3.2 Tahap Pengerjaan Penelitian . . . . .	21
3.3 Data Cleaning . . . . .	22
3.4 Data Integration . . . . .	24
3.5 Data Selection . . . . .	25
3.6 Data Transformation . . . . .	26
3.6.1 Normalization . . . . .	26
3.6.2 Binning . . . . .	27
3.7 Linear Regression . . . . .	28
3.8 Multiple Linear Regression . . . . .	30
3.9 Evaluasi Hasil Regresi . . . . .	32
3.10 Decision Tree(ID3) . . . . .	32

3.11 Naïve Bayes . . . . .	44
3.12 Evaluasi Klasifikasi . . . . .	45
3.13 k-Means . . . . .	47
3.14 Agglomerative Nesting(AGNES) . . . . .	49
3.15 Evaluasi Clustering . . . . .	52
<b>4 ANALISIS DAN IMPLEMENTASI</b>	<b>55</b>
4.1 Lingkungan Perangkat Keras . . . . .	55
4.2 Lingkungan Perangkat Lunak . . . . .	55
4.3 Deskripsi Dataset . . . . .	55
4.4 Proses Analisis Data . . . . .	57
4.4.1 Penyiapan Data . . . . .	58
4.4.2 Analisis Data . . . . .	71
4.4.2.1 Analisis dengan Teknik Regresi . . . . .	71
4.4.2.2 Analisis dengan Teknik Klasifikasi . . . . .	76
4.4.2.3 Analisis dengan Teknik Clustering . . . . .	84
4.5 Proses Analisis Data Non Agregat . . . . .	91
4.5.1 Penyiapan Data . . . . .	91
4.5.2 Analisis Data . . . . .	95
4.5.2.1 Analisis dengan Teknik Regresi . . . . .	95
4.5.2.2 Analisis dengan Teknik Klasifikasi . . . . .	100
4.5.2.3 Analisis dengan Teknik Clustering . . . . .	113
4.6 Pembahasan Hasil Analisis . . . . .	125
<b>5 KESIMPULAN DAN SARAN</b>	<b>129</b>
5.1 Kesimpulan . . . . .	129
5.2 Saran . . . . .	130
<b>DAFTAR REFERENSI</b>	<b>131</b>
<b>A KODE PROGRAM</b>	<b>133</b>

## DAFTAR GAMBAR

2.1 Tahapan Wirausaha . . . . .	6
2.2 Knowledge Dsicovery . . . . .	9
2.3 Ilustrasi Algoritma <i>Clustering</i> k-Means . . . . .	17
2.4 Contoh Agglo . . . . .	18
3.1 Decision Tree Tahap 1 . . . . .	37
3.2 Decision Tree Tahap 1 Final . . . . .	37
3.3 Decision Tree Tahap 2 . . . . .	40
3.4 Decision Tree Tahap 3 . . . . .	42
3.5 Decision Tree Tahap 3 . . . . .	43
3.6 Gambar Agglo . . . . .	50
4.1 Kolom NA APS . . . . .	59
4.2 Kolom NA NES . . . . .	59
4.3 Distribusi Perceived Opportunities . . . . .	61
4.4 Distribusi TEA . . . . .	62
4.5 Distribusi EBO . . . . .	62
4.6 Distribusi Cultural and Social Norms . . . . .	63
4.7 Distribusi Basic Entrepreneurial Education and Training . . . . .	64
4.8 Distribusi Taxes and Bureaucracy . . . . .	64
4.9 Pearson Correlation APS . . . . .	66
4.10 Pearson Correlation NES . . . . .	67
4.11 Scatter Plot Perceived Opportunities dengan Perceived Capabilities . . . . .	71
4.12 Scatter Plot Entrepreneurial Employee Activity dengan Business Services Sector . . . . .	72
4.13 Scatter Plot Perceived Capabilities dengan Entrepreneurship as a Good Career Choice . . . . .	72
4.14 Scatter Plot Governmental support and policies dengan Governmental programs . . . . .	73
4.15 Scatter Plot Basic school entrepreneurial education and training dengan Post school entrepreneurial education and training . . . . .	73
4.16 Scatter Plot Financing for entrepreneurs dengan R&D transfer . . . . .	74
4.17 Decision Tree Perceived Opportunities . . . . .	77
4.18 Decision Tree Perceived Capabilities . . . . .	78
4.19 Decision Tree Entrepreneurial intentions . . . . .	78
4.20 Decision Tree Entrepreneurial Employee Activity . . . . .	79
4.21 Decision Tree Financing for entrepreneurs . . . . .	80
4.22 Decision Tree Governmental support and policies . . . . .	81
4.23 Decision Tree Taxes and bureaucracy . . . . .	81
4.24 Decision Tree Governmental programs . . . . .	82
4.25 Elbow Method untuk Total early-stage Entrepreneurial Activity (TEA) dan Established Business Ownership . . . . .	84
4.26 Cluster yang terbentuk dari Total early-stage Entrepreneurial Activity (TEA) dan Established Business Ownership dengan teknik k-Means . . . . .	85
4.27 Elbow Method untuk Governmental support and policies dan Governmental programs . . . . .	86

4.28 Cluster yang terbentuk dari Governmental support and policies dan Governmental programs dengan teknik k-Means . . . . .	87
4.29 Dendrogram yang terbentuk dari Governmental support and policies dan Governmental programs dengan teknik AGNES . . . . .	88
4.30 Dendrogram yang terbentuk dari Total early-stage Entrepreneurial Activity (TEA) dan Established Business Ownership dengan teknik AGNES . . . . .	89
4.31 Pearson Correlation APS non Agregat . . . . .	92
4.32 Pearson Correlation NES non Agregat . . . . .	93
4.33 Scatter Plot Perceived Opportunities dengan Perceived Capabilities untuk data Non Agregat . . . . .	95
4.34 Scatter Plot Entrepreneurial Employee Activity dengan Business Services Sector untuk data non Agregat . . . . .	96
4.35 Scatter Plot Perceived Capabilities dengan Entrepreneurship as a Good Career Choice untuk data non Agregat . . . . .	96
4.36 Scatter Plot Governmental support and policies dengan Governmental programs untuk data non Agregat . . . . .	97
4.37 Scatter Plot Basic school entrepreneurial education and training dengan Post school entrepreneurial education and training non Agregat . . . . .	97
4.38 Scatter Plot Financing for entrepreneurs dengan R&D transfer untuk data non Agregat . . . . .	98
4.39 Decision Tree Perceived Opportunities untuk data non Agregat . . . . .	101
4.40 Decision Tree Perceived Capabilities untuk data non Agregat . . . . .	102
4.41 Decision Tree Entrepreneurial intentions untuk data non Agregat . . . . .	103
4.42 Decision Tree Entrepreneurial Employee Activity untuk data non Agregat . . . . .	104
4.43 Decision Tree Financing for entrepreneurs untuk data non Agregat . . . . .	107
4.44 Decision Tree Governmental support and policies untuk data non Agregat . . . . .	108
4.45 Decision Tree Taxes and bureaucracy untuk data non Agregat . . . . .	109
4.46 Decision Tree Governmental programs untuk data non Agregat . . . . .	110
4.47 Elbow Method untuk Total early-stage Entrepreneurial Activity (TEA) dan Established Business Ownership untuk data Non Agregat . . . . .	113
4.48 Cluster yang terbentuk dari Total early-stage Entrepreneurial Activity (TEA) dan Established Business Ownership dengan teknik k-Means untuk data Non Agregat . . . . .	114
4.49 Elbow Method untuk Governmental support and policies dan Governmental programs untuk data Non Agregat . . . . .	116
4.50 Cluster yang terbentuk dari Governmental support and policies dan Governmental programs dengan teknik k-Means untuk data Non Agregat . . . . .	117
4.51 Dendrogram yang terbentuk dari Total early-stage Entrepreneurial Activity (TEA) dan Established Business Ownership dengan teknik AGNES untuk data Non Agregat . . . . .	120
4.52 Dendrogram yang terbentuk dari Total early-stage Entrepreneurial Activity (TEA) dan Established Business Ownership dengan teknik AGNES untuk data Non Agregat . . . . .	124

## DAFTAR TABEL

1.1	Contoh indikator APS untuk Indonesia dan Malaysia pada tahun 2016 dan 2017 . . . . .	1
1.2	Contoh indikator NES untuk Indonesia dan Malaysia pada tahun 2016 dan 2017 . . . . .	1
2.1	Confusion Matrix . . . . .	15
3.1	Data kotor . . . . .	22
3.2	Penyanyi . . . . .	24
3.3	Judul . . . . .	24
3.4	Penyanyi dan Judul yang sudah tergabung . . . . .	24
3.5	Tinggi Badan . . . . .	25
3.6	Korelasi Tinggi Badan . . . . .	25
3.7	Atribut Sebelum Normalisasi . . . . .	26
3.8	Atribut Sesudah Normalisasi . . . . .	26
3.9	Atribut Sebelum Binning . . . . .	27
3.10	Atribut Sesudah Binning . . . . .	27
3.11	Atribut Regresi . . . . .	28
3.12	Perhitungan Konstanta Linear Regression . . . . .	28
3.13	Hasil Perhitungan Linear Regression . . . . .	29
3.14	Atribut Multiple Linear Regression . . . . .	30
3.15	Perhitungan Konstanta Multiple Linear Regression . . . . .	30
3.16	Hasil Perhitungan Multiple Linear Regression . . . . .	31
3.17	Perhitungan Evaluasi Multiple Linear Regression . . . . .	32
3.18	Atribut Decision Tree . . . . .	33
3.19	Atribut Perhitungan Naive Bayes . . . . .	44
3.20	Atribut Evaluasi Klasifikasi . . . . .	46
3.21	Atribut k-Means . . . . .	47
3.22	Perhitungan k-Means iterasi 1 . . . . .	47
3.23	Perhitungan k-Means iterasi 2 . . . . .	47
3.24	Perhitungan k-Means iterasi 2 . . . . .	48
3.25	Perhitungan k-Means iterasi 3 . . . . .	48
3.26	Perhitungan k-Means iterasi 3 . . . . .	48
3.27	Atribut Agglomerative Nesting . . . . .	49
3.28	Perhitungan Euclian Distance iterasi 1 . . . . .	49
3.29	Perhitungan Euclian Distance iterasi 2 . . . . .	50
3.30	Perhitungan Euclian Distance iterasi 3 . . . . .	50
3.31	Perhitungan Euclian Distance iterasi 4 . . . . .	50
3.32	Evaluasi Clustering . . . . .	52
4.1	Deskripsi Dataset APS . . . . .	56
4.2	Contoh Atribut APS . . . . .	56
4.3	Deskripsi Dataset NES . . . . .	57
4.4	Contoh Atribut NES . . . . .	57
4.5	Peserta APS pada 2014-2021 . . . . .	58

4.6	Peserta NES Pada 2014-2021	60
4.7	Contoh Atribut APS Non Agregat	60
4.8	Contoh Atribut NES Non Agregat	61
4.9	Contoh Atribut APS Agregat	63
4.10	Contoh Atribut NES Agregat	65
4.11	Korelasi APS yang $\geq 0.65$	67
4.12	Korelasi NES yang $\geq 0.65$	68
4.13	Contoh Atribut APS yang telah di Binning	69
4.14	Contoh Atribut NES yang telah di Binning	70
4.15	Koefisien Determinasi atribut APS Agregat	76
4.16	Koefisien Determinasi atribut NES Agregat	76
4.17	Accuracy, Precision, Recall Decision Tree atribut APS	77
4.18	Accuracy, Precision, Recall Decision Tree atribut NES	81
4.19	Accuracy, Precision, Recall Naïve Bayes atribut APS	83
4.20	Accuracy, Precision, Recall Naïve Bayes atribut NES	83
4.21	Hasil Clustering atribut APS dengan k-Means	85
4.22	Hasil Clustering atribut NES dengan k-Means	87
4.23	Kode Negara	90
4.24	Korelasi APS untuk data Non Agregat	92
4.25	Korelasi NES untuk data Non Agregat	94
4.26	Koefisien Determinasi atribut APS non Agregat	100
4.27	Koefisien Determinasi atribut NES non Agregat	100
4.28	Accuracy, Precision, Recall Decision Tree atribut APS untuk data non Agregat	105
4.29	Accuracy, Precision, Recall Decision Tree atribut NES untuk data non Agregat	111
4.30	Accuracy, Precision, Recall Naïve Bayes atribut APS untuk data non Agregat	112
4.31	Accuracy, Precision, Recall Naïve Bayes atribut NES untuk data non Agregat	112
4.32	Hasil Clustering atribut APS non agregat dengan k-Means	115
4.33	Hasil Clustering atribut NES non agregat dengan k-Means	118
4.34	Kode negara untuk atribut APS non agregat	122
4.35	Kode negara untuk atribut NES non agregat	123

## DAFTAR KODE PROGRAM

3.1 <i>Data Cleaning</i> . . . . .	22
3.2 Operasi <i>Join</i> . . . . .	24
3.3 Nilai Pearson <i>Correlation</i> antar fitur . . . . .	25
3.4 <i>Min-Max Normalization</i> . . . . .	26
3.5 <i>Binning</i> . . . . .	27
3.6 <i>Linear Regression</i> . . . . .	29
3.7 <i>Multiple Linear Regression</i> . . . . .	31
3.8 Komputasi Koefisien Determinasi . . . . .	32
3.9 Pembuatan <i>Decision Tree</i> . . . . .	43
3.10 Komputasi Naïve Bayes . . . . .	45
3.11 Evaluasi model Klasifikasi . . . . .	46
3.12 Komputasi <i>k-Means</i> . . . . .	49
3.13 Komputasi <i>Agglomerative</i> . . . . .	51
3.14 Komputasi <i>Silhoutte Coefficient</i> . . . . .	53
A.1 <i>Data Cleaning</i> . . . . .	133
A.2 Operasi <i>Join</i> . . . . .	133
A.3 Nilai Pearson <i>Correlation</i> antar fitur . . . . .	134
A.4 <i>Min-Max Normalization</i> . . . . .	134
A.5 <i>Binning</i> . . . . .	134
A.6 <i>Linear Regression</i> . . . . .	134
A.7 <i>Multiple Linear Regression</i> . . . . .	134
A.8 Komputasi Koefisien Determinasi . . . . .	135
A.9 Pembuatan <i>Decision Tree</i> . . . . .	135
A.10 Komputasi Naïve Bayes . . . . .	135
A.11 Evaluasi model Klasifikasi . . . . .	136
A.12 Komputasi <i>k-Means</i> . . . . .	136
A.13 Komputasi <i>Agglomerative</i> . . . . .	136
A.14 Komputasi <i>Silhoutte Coefficient</i> . . . . .	137
A.15 Analisis data APS Agregat . . . . .	137
A.16 Analisis data APS Non Agregat . . . . .	144
A.17 Analisis data NES Agregat . . . . .	152
A.18 Analisis data NES Non Agregat . . . . .	159





# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Kegiatan wirausaha memegang peranan penting dalam pembangunan suatu negara atau dapat dikatakan bahwa kewirausahaan salah satu penggerak roda ekonomi suatu negara<sup>1</sup>. Karena orang-orang mulai sadar betapa pentingnya kegiatan wirausaha disuatu negara, maka pada tahun September 1997 *Global Entrepreneurship Monitor*(GEM) didirikan oleh Michael Hay dari London Business School(LBS) dan Bill Bygrave dari Babson College[1].

*Global Entrepreneurship Monitor*(GEM) merupakan sebuah organisasi non-profit yang berfokus pada kewirausahaan. Pada tahun 2021, GEM mempunyai 115 negara sebagai anggotanya. Setiap tahunnya GEM membuat sebuah laporan terkait kewirausahaan di suatu negara. Laporan tersebut dibuat berdasarkan dua buah survei yang diadakan oleh GEM. Survei tersebut adalah *Adult Population Survey*(APS) dan *National Expert Survey*(NES)[1]. APS merupakan survei yang dilakukan dengan setidaknya 2000 orang dewasa yang berusia 18 hingga 64 tahun yang dipilih secara acak untuk setiap negara. Tujuan survei APS adalah untuk melihat karakteristik, motivasi, dan ambisi individu yang memulai bisnis, serta sikap sosial terhadap kewirausahaan<sup>2</sup>. Berikut merupakan beberapa contoh indikator yang menjadi penilaian dalam survei APS<sup>2</sup>.

Tabel 1.1: Contoh indikator APS untuk Indonesia dan Malaysia pada tahun 2016 dan 2017

Negara	Tahun	<i>Perceived Opportunities</i>	<i>Perceived Capabilities</i>	<i>Fear of Failure Rate</i>
Indonesia	2016	43.10	55.10	38.80
Indonesia	2017	47.10	57.30	46.70
Malaysia	2016	25.40	28.30	36.70
Malaysia	2017	45.10	46.10	45.00

NES merupakan survei kepada pakar yang bertujuan untuk memberikan penilaian dan wawasan secara mendalam tentang iklim kewirausahaan di masing-masing negara<sup>3</sup>. Berikut merupakan beberapa contoh indikator yang menjadi penilaian dalam survei NES<sup>2</sup>.

Tabel 1.2: Contoh indikator NES untuk Indonesia dan Malaysia pada tahun 2016 dan 2017

Negara	Tahun	<i>Financing For Entrepreneurs</i>	<i>Governmental Policies: Support And Relevance</i>	<i>Governmental Policies: Taxes And Bureaucracy</i>
Indonesia	2016	2.77	2.77	2.27
Indonesia	2017	3.71	3.74	3.42
Malaysia	2016	3.20	2.98	2.57
Malaysia	2017	3.48	2.74	2.50

<sup>1</sup><https://www.kompasiana.com/riki1987/551fd72f813311f3379df4df/kewirausahaan-di-indonesia-mempengaruhi-perekonomian>

<sup>2</sup><https://www.gemconsortium.org/wiki/1154>

<sup>3</sup><https://www.gemconsortium.org/wiki/1142>

Data yang dihasilkan dalam survei APS dan NES, dapat dimanfaatkan jika kita dapat mencari *insight* yang tersembunyi di dalam data tersebut. Pada Tabel 1.1 pada kolom *Perceived Opportunities* terlihat bahwa penduduk Malaysia pada tahun 2017 lebih optimis akan memulai bisnis dibandingkan pada tahun 2016 atau pada Tabel 1.2 pada kolom *Governmental Policies: Support And Relevance* berdasarkan pendapat ahli terlihat bahwa pada tahun 2017 pemerintah Indonesia membuat kebijakan publik yang lebih mendukung kegiatan wirausaha sebagai isu ekonomi yang relevan dibandingkan pada tahun 2016. Data yang ada juga dapat bermanfaat bagi banyak pihak misalnya untuk para akademisi, sponsor kegiatan survei, wirausahawan, maupun organisasi internasional. Atribut pada Tabel 1.1 dan Tabel 1.2 yang bentuknya numerik bisa dimanfaatkan juga untuk melakukan prediksi nilai-nilai pada indikator lainnya menggunakan teknik regresi, dimana regresi merupakan salah satu teknik *Data Mining*.

*Data Mining* merupakan sebuah teknik untuk mengambil informasi dari sekumpulan data yang jumlahnya besar[2]. *Data Mining* memiliki beberapa teknik, namun yang akan digunakan pada skripsi ini adalah regresi, klasifikasi, dan *clustering*. Dengan menggunakan teknik prediktif, data GEM untuk tiap negara akan dimanfaatkan untuk memprediksi atribut suatu label berdasarkan atribut yang ada pada hasil survei APS dan NES. Kemudian, dengan menggunakan teknik *clustering*, data GEM untuk tiap negara akan dikelompokkan berdasarkan suatu atribut pada hasil survei APS dan NES.

Berdasarkan penelitian sejenis yang menggunakan data GEM, ada artikel yang berjudul *Determinants of Growth Aspirations: Empirical Evidence of the South-Eastern European Countries*[3]. Artikel ini menghasilkan *insight* menarik namun hanya untuk daerah Eropa Tenggara. Pada skripsi ini, akan dibuat sebuah perangkat lunak yang dapat menampilkan visualisasi dan hasil analisis data GEM untuk wilayah Asia. Perangkat lunak yang akan dibangun juga dapat melakukan tahapan *Data Mining* seperti membersihkan data dan menjalankan algoritma *Data Mining*.

## 1.2 Rumusan Masalah

Berikut ini adalah rumusan masalah berdasarkan latar belakang yang telah dipaparkan:

1. Apa itu data GEM?
2. Teknik-teknik *Data Mining* apa yang dapat digunakan untuk mengetahui keterkaitan antar atribut, mengelompokkan data berdasarkan kelas tertentu, dan kemiripan tertentu?
3. Bagaimana cara untuk memprediksi atribut numerik menggunakan data APS dan NES?
4. Bagaimana cara memetakan atribut APS dan NES ke dalam label tertentu?
5. Bagaimana cara melihat kelompok yang terbentuk berdasarkan kemiripan atribut APS dan NES?
6. Bagaimana cara menyiapkan data agar data siap dianalisis?

## 1.3 Tujuan

Tujuan yang ingin dicapai dari penelitian ini adalah:

1. Mempelajari apa itu data GEM serta mempelajari data-data yang digunakan dalam survei APS dan NES.
2. Mempelajari teknik *Data Mining* untuk menganalisis data GEM diantaranya teknik regresi, klasifikasi, dan *clustering*.
3. Menggunakan teknik *Data Mining* untuk menganalisis data GEM yaitu teknik regresi.
4. Menggunakan teknik *Data Mining* untuk menganalisis data GEM yaitu klasifikasi.
5. Menggunakan teknik *Data Mining* untuk menganalisis data GEM yaitu *clustering*.
6. Membangun perangkat lunak yang mampu melakukan persiapan dan analisis data GEM.

## 1.4 Batasan Masalah

Pelaksanaan penelitian ini permasalahannya dibatasi pada:

1. *Dataset* yang digunakan berasal dari situs resmi GEM untuk negara-negara di Asia.
2. *Dataset* yang digunakan merupakan data APS dan NES dari tahun 2014 sampai 2021.
3. Penelitian ini akan menggunakan *library* Python untuk mengimplementasikan tahapan *Data Mining*.
4. Perangkat lunak yang dibuat tidak memiliki tampilan antar muka.

## 1.5 Metodologi

Langkah-langkah yang akan dilakukan dalam melakukan penelitian ini, yaitu:

1. Melakukan studi literatur mengenai GEM, serta indikator-indikator apa saja yang digunakan dalam survei APS dan NES, prapengolahan data, jenis data, dan teknik-teknik *Data Mining*.
2. Menerapkan tahapan prapengolahan data dan teknik-teknik *Data Mining* pada data APS dan NES meliputi teknik Regresi, Klasifikasi, dan *Clustering*.
3. Melakukan analisis dan menuliskan kesimpulan apa saja yang telah didapatkan dari data APS dan NES.

## 1.6 Sistematika Pembahasan

Sistematika penulisan ini berguna untuk memberikan gambaran secara umum mengenai penelitian yang akan dibuat. Berikut ini adalah uraian dari sistematika pembahasan:

- Bab 1. Pendahuluan, latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, metode penelitian, dan sistematika pembahasan skripsi.
- Bab 2. Landasan Teori, membahas tentang GEM, pengertian *Data Mining* serta teknik-tekniknya, WordPress, Python, dan Tableau.
- Bab 3. Studi Eksplorasi Algoritma Data Mining, membahas bagaimana implementasi algoritma yang telah dijelaskan pada bab sebelumnya.
- Bab 4. Analisis dan Implementasi, membahas tentang deskripsi *dataset*, hasil analisis data dengan teknik *Data Mining*.
- Bab 5. Kesimpulan dan Saran, membahas kesimpulan dari eksperimen yang telah dilakukan dan saran untuk penelitian selanjutnya.

