

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada hadirat Tuhan Yang Maha Esa karena dengan rahmat dan izin-Nya penulis dapat menyelesaikan skripsi dengan baik dan tepat waktu.

Skripsi dengan judul “IMPLEMENTASI *MACHINE LEARNING* UNTUK ANALISIS KLASSTER DAN ANALISIS PREDIKTIF DATA BPJS KESEHATAN INDONESIA TAHUN 2015-2016” ini buat dengan tujuan sebagai persyaratan untuk memperoleh gelar Sarjana Teknik Strata Satu di Universitas Pelita Harapan, Tangerang.

Penelitian dan penulisan skripsi sangatlah mustahil jika dikerjakan sendirian dan tanpa ada dukungan. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada pihak – pihak yang terlibat pada proses penyelesaian skripsi ini, yaitu :

1. Bapak Eric Jobiliong, Ph.D. selaku Dekan Fakultas Sains dan Teknologi Universitas Pelita Harapan.
2. Bapak Dr. Henri P. Uranus selaku Ketua Program Studi Teknik Elektro Universitas Pelita Harapan.
3. Bapak Dr.-Ing. Ihan Martoyo, MTS selaku pembimbing utama yang telah memberikan begitu banyak wawasan, motivasi, dan bimbingan selama penyusunan skripsi ini.
4. Bapak Kie Van Ivanky S., S.Si., Ph.D. selaku pembimbing pendamping yang telah memberikan begitu banyak wawasan, motivasi dan bimbingan selama penyusunan skripsi ini.
5. Pihak BPJS Kesehatan Indonesia yang telah menyediakan data untuk dapat diolah dalam skripsi ini.
6. Keluarga dari penulis yang telah memberikan motivasi dan dukungan untuk penulis.
7. Seluruh rekan mahasiswa Teknik Elektro angkatan 2017 yang telah menjadi teman seperjuangan sejak awal perkuliahan hingga akhir, terutama kepada Timothius Hermawan dan Hans Reinaldy yang sama – sama berjuang dengan topik penelitian serupa.

8. Seluruh dosen, laboran, dan staf Universitas Pelita Harapan terutama di Program Studi Teknik Elektro yang telah memberikan ilmu dan pengetahuan selama penulis berkuliah di Universitas Pelita Harapan.
9. Seluruh rekan mahasiswa Teknik Elektro dari angkatan 2014 sampai 2019, terutama kepada Gerry Chandra, Reynold V. Chen, Alvaro Amos, dan Edward Samudra yang telah membantu dan membimbing penulis sebelum pengerjaan skripsi ini dimulai sehingga ilmu tersebut dapat digunakan sekarang.
10. Pihak – pihak lain yang tidak dapat disebutkan satu per satu.

Akhir kata, penulis sangat mengharapkan kritik dan saran yang membangun dari seluruh pembaca agar skripsi yang jauh dari kata sempurna ini dapat menjadi lebih baik dan lebih bermanfaat.

Tangerang, 18 Februari 2021

(Winly Williamdy)

DAFTAR ISI

halaman

HALAMAN JUDUL	
PERNYATAAN DAN PERSETUJUAN UNGGAH TUGAS AKHIR	
PERSETUJUAN DOSEN PEMBIMBING SKRIPSI	
PERSETUJUAN TIM PENGUJI SKRIPSI	
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL.....	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Maksud dan Tujuan.....	3
1.3. Batasan Masalah.....	4
1.4. Metode Penelitian.....	5
1.5. Sistematika Penelitian	5
BAB II LANDASAN TEORI	7
2.1. <i>Data Science</i>	7
2.2. <i>Descriptive Analytics</i>	7
2.3. <i>Predictive Analytics</i>	8
2.4. <i>Python</i>	8
2.5. <i>Machine Learning</i>	9
2.6. <i>Unsupervised Learning</i>	9
2.6.1. Analisis <i>RFM</i>	10
2.6.2. <i>Clustering</i>	11
2.6.3. <i>Elbow Method</i>	15
2.7. <i>Supervised Learning</i>	17
2.7.1. Klasifikasi	17
2.7.2. Regresi	19
2.7.3. <i>Time Series Forecasting</i>	20
2.7.4. Teori Kredibilitas / <i>Credibility Theory</i>	24
2.8. Metrik Penilaian Model.....	26
2.8.1. Metrik Penilaian Model Regresi & <i>Time Series Forecasting</i>	27
2.9. Sistem Penyelenggaraan Jaminan Kesehatan Indonesia	28
2.9.1. Segmentasi Peserta pada BPJS Kesehatan	28
2.9.2. Fasilitas Kesehatan dalam Sistem BPJS Kesehatan Indonesia...30	
2.9.3. Sistem Pendanaan Fasilitas Kesehatan Tingkat Pertama.....32	
2.10. Perubahan Kebijakan BPJS Kesehatan Selama Tahun 2015 - 2016.33	
2.11. Sistem BPJS Kesehatan Tahun 2020	34

BAB III METODE PENELITIAN	35
3.1. Deskripsi Sistem.....	35
3.1.1. Deskripsi Perangkat	35
3.1.2. Deskripsi <i>Datasets</i>	35
3.2. <i>Flowchart</i> Penelitian	44
3.2.1. Integrasi <i>Dataset</i>	45
3.2.2. <i>Data Cleansing</i>	46
3.2.3. Perhitungan <i>RFM</i> untuk Setiap Peserta	48
3.2.4. <i>Exploratory Data Analysis</i>	51
3.2.5. Standarisasi Data	51
3.2.6. <i>Elbow Method</i>	52
3.2.7. Klasterisasi dengan Algoritma <i>K-Means Clustering</i>	52
3.2.8. Visualisasi dan Deskripsi Klaster	52
3.2.9. Pembentukan Model Prediksi	53
 BAB IV <i>EXPLORATORY DATA ANALYSIS</i>	 55
4.1. <i>Dataset</i> Kepesertaan.....	55
4.1.1. Jenis Kelamin.....	55
4.1.2. Kategori Umur	56
4.1.3. Segmen Peserta.....	56
4.1.4. Status Perkawinan.....	57
4.1.5. Kelas Rawat	58
4.2. <i>Dataset</i> Kunjungan.....	58
4.2.1. Diagnosis Terbanyak	59
4.2.2. Jenis Kelamin.....	60
4.2.3. Kategori Umur	61
4.2.4. Segmen Peserta.....	62
4.2.5. Status Perkawinan.....	63
4.2.6. Kelas Rawat	63
4.2.7. Status Pulang	64
4.2.8. <i>Trend</i> Kunjungan	65
4.3. <i>Dataframe RFM</i>	67
4.4. <i>Dataframe</i> Kunjungan per Bulan	69
 BAB V HASIL DAN PEMBAHASAN.....	 70
5.1 Hasil Klasterisasi.....	70
5.1.1. <i>Elbow Method</i> dan Jumlah Iterasi.....	70
5.1.2. Deskripsi Klaster	72
5.2. Hasil Prediksi	79
5.2.1. <i>One Step Forecasting</i>	80
5.2.1.1. <i>Bühlmann-Straub Credibility Theory</i>	80
5.2.1.2. <i>Random Forest Regressor</i>	84
5.2.1.2. <i>Linear Regression</i>	86
5.2.1.4. <i>ARIMA</i>	90
5.2.2. <i>Multi Step Forecasting</i>	95
5.2.2.1. <i>Random Forest Regressor</i>	95

5.2.1.1. <i>Linear Regression</i>	97
5.2.1.3. <i>ARIMA</i>	101
BAB VI KESIMPULAN DAN SARAN	106
6.1. Kesimpulan.....	106
6.2. Saran.....	110
DAFTAR PUSTAKA	114
LAMPIRAN	116



DAFTAR GAMBAR

	halaman
Gambar 2.1	Ilustrasi <i>Euclidean distance</i> pada dua dimensi dan rumus perhitungan <i>Euclidean distance</i> untuk n dimensi11
Gambar 2.2	(a) Hasil pengelompokan dengan menggunakan <i>K-Means Clustering</i> dengan <i>dataset</i> yang memiliki sebaran unik11 (b) Hasil pengelompokan dengan menggunakan <i>Spectral Clustering</i> dengan <i>dataset</i> yang memiliki sebaran unik.12
Gambar 2.3	(a) Awal iterasi dari <i>K-Means Clustering</i>13 (b) Pergeseran <i>centroid</i> menuju <i>mean</i> setiap klaster.....13 (c) Perhitungan jarak untuk posisi <i>centroid</i> yang baru13 (d) Sebaran data setelah algoritma <i>K-Means Clustering</i> selesai13
Gambar 2.4	Hasil perhitungan jarak antara sampel A dan <i>centroid</i> pada iterasi pertama menggunakan <i>Euclidean distance</i>14
Gambar 2.5	<i>Elbow method</i> untuk menentukan nilai K yang sesuai.....15
Gambar 2.6	<i>Least square method</i> untuk menentukan persamaan <i>linear regression</i>19
Gambar 3.1	<i>Flowchart</i> penelitian43
Gambar 3.2	Ilustrasi <i>dataset</i> yang digunakan dalam penelitian44
Gambar 4.1	Sebaran jenis kelamin dalam <i>dataset</i> Kepesertaan54
Gambar 4.2	Sebaran kategori umur dalam <i>dataset</i> Kepesertaan55
Gambar 4.3	Sebaran segmentasi peserta dalam <i>dataset</i> Kepesertaan56
Gambar 4.4	Sebaran status perkawinan dalam <i>dataset</i> Kepesertaan56
Gambar 4.5	Sebaran kelas rawat dalam <i>dataset</i> Kepesertaan57
Gambar 4.6	Sepuluh diagnosis terbanyak pada data Kunjungan FKTP58
Gambar 4.7	Sepuluh diagnosis terbanyak pada data Kunjungan FKRTL59
Gambar 4.8	Perbandingan sebaran jenis kelamin pada data Kunjungan FKTP dan FKRTL60
Gambar 4.9	Sebaran kategori umur dalam data Kunjungan FKTP dan FKRTL60
Gambar 4.10	Sebaran segmentasi peserta dalam data Kunjungan FKTP dan FKRTL61
Gambar 4.11	Sebaran status perkawinan dalam data Kunjungan FKTP dan FKRTL62
Gambar 4.12	Sebaran kelas rawat dalam data Kunjungan FKTP dan FKRTL63
Gambar 4.13	Sebaran status pulang dalam data Kunjungan FKRTL64
Gambar 4.14	Sebaran hari berkunjung dalam data Kunjungan FKTP dan FKRTL65
Gambar 4.15	Trend penambahan jumlah Kunjungan di FKTP dan FKRTL selama dua tahun65
Gambar 4.16	Trend penambahan persentase Kunjungan pada FKTP dan FKRTL selama dua tahun66
Gambar 5.1	Grafik distorsi setiap klaster K dari K = 2 hingga K = 1469

Gambar 5.2	Grafik jumlah iterasi untuk setiap nilai K.....	71
Gambar 5.3	Hasil klasterisasi menggunakan algoritma <i>K-Means Clustering</i> dengan $K = 4$	72
Gambar 5.4	Sebaran jenis kelamin peserta pada setiap klaster.....	75
Gambar 5.5	Persebaran segmentasi umur peserta pada setiap klaster	76
Gambar 5.6	Sebaran segmentasi peserta pada setiap klaster	77
Gambar 5.7	Sebaran status perkawinan pada setiap klaster.....	78
Gambar 5.8	<i>Feature importance</i> dari model <i>Random Forest Regressor</i> pada skenario <i>one step forecasting</i>	84
Gambar 5.9	Korelasi antara setiap variabel prediktor dengan variabel target.....	85
Gambar 5.10	Distribusi <i>residual</i> model <i>Linear Regression</i> pada skenario <i>one step forecasting</i>	88
Gambar 5.11	(a) Visualisasi <i>Partial AutoCorrelation Function</i>	90
	(b) Visualisasi <i>AutoCorrelation Function</i>	90
Gambar 5.12	<i>Feature importance</i> model <i>Random Forest Regressor</i> pada skenario <i>multi step forecasting</i>	95
Gambar 5.13	Hasil prediksi model <i>Random Forest Regressor</i> pada skenario <i>multi step forecasting</i>	95
Gambar 5.14	Sedaran <i>residual</i> model <i>Linear Regression</i> pada skenario <i>multi step regression</i>	98
Gambar 5.15	Hasil prediksi menggunakan <i>Linear Regression</i> pada skenario <i>multi step forecasting</i>	98
Gambar 5.16	Hasil prediksi model <i>Linear Regression</i> dengan <i>feature importance</i> sebagai <i>predictor</i>	99
Gambar 5.17	Hasil prediksi menggunakan <i>ARIMA</i> pada skenario <i>multi step forecasting</i>	101
Gambar 5.18	Perbandingan hasil prediksi setiap model untuk skenario <i>multi step forecasting</i>	102

DAFTAR TABEL

	halaman
Tabel 2.1	Perubahan iuran segmen non-PBI pada Perpres no.19 tahun 2016.....32
Tabel 2.2	Perubahan iuran BPJS Kesehatan tahun 202033
Tabel 3.1	Versi perangkat dan <i>libraries</i>34
Tabel 3.2	Rincian kolom pada <i>dataset</i> Kepesertaan.35
Tabel 3.3	Rincian kolom pada <i>dataset</i> FKTP Nonkapitasi.....37
Tabel 3.4	Rincian kolom pada <i>dataset</i> FKRTL39
Tabel 3.5	Rincian kolom pada <i>dataframe</i> RFM.....48
Tabel 3.6	Iuran Segmen Non-PBI49
Tabel 4.1	(a) Ringkasan statistik untuk <i>dataframe</i> RFM dengan peserta yang tidak pernah berkunjung di FKTP dan FKRTL.....67
	(b) Ringkasan statistik untuk <i>dataframe</i> RFM tanpa peserta yang tidak pernah berkunjung di FKTP dan FKRTL.....67
Tabel 5.1	Distorsi setiap klaster dan penurunannya.....70
Tabel 5.2	Ringkasan statistik setiap klaster73
Tabel 5.3	Jumlah kunjungan sesuai periode waktu dan kelompok peserta81
Tabel 5.4	<i>Credibility factor</i> dan nilai rata – rata kelompok untuk setiap kelompok peserta82
Tabel 5.5	Hasil prediksi menggunakan model <i>Bühlmann-Straub Credibility Theory</i>83
Tabel 5.6	Ringkasan performa model <i>Linear Regression</i> pada skenario <i>one step forecasting</i>84
Tabel 5.7	Tabel ringkasan model untuk model <i>Linear Regression</i> menggunakan <i>feature importance</i>88
Tabel 5.8	Hasil uji stasioneritas <i>Augmented Dickey-Fuller</i>89
Tabel 5.9	Tabel ringkasan model ARIMA pada skenario <i>one step forecasting</i>92
Tabel 5.10	Perbandingan hasil prediksi berbagai model pada skenario <i>one step forecasting</i>93
Tabel 5.11	Ringkasan performa model <i>Linear Regression</i> pada skenario <i>multi step forecasting</i>96
Tabel 5.12	Ringkasan model <i>Linear Regression</i> dengan <i>feature importance</i> sebagai prediktor100
Tabel 5.13	Tabel ringkasan model ARIMA pada skenario <i>multi step forecasting</i>101
Tabel 5.14	Perbandingan <i>RMSE</i> ketiga model dalam skenario <i>multi step forecasting</i>103
Tabel 6.1	Rangkuman karakteristik klaster berdasarkan jumlah penghuni klaster pada beberapa variabel106

DAFTAR LAMPIRAN

	halaman
LAMPIRAN A :	
Kode Program : <i>EDA & Clustering</i>	A-1
LAMPIRAN B :	
Kode Program : <i>Visualization</i>	B-1
LAMPIRAN C :	
Kode Program : <i>Prediction</i>	C-1
LAMPIRAN D :	
Lampiran B-1 : Lembar Bimbingan (Pembimbing Utama)	D-1
Lampiran B-2 : Lembar Bimbingan (Pembimbing Pendamping)	D-2
LAMPIRAN E :	
Lampiran E-1 : Form Hasil Uji Similaritas	E-1
Lampiran E-2 : Hasil Uji Similaritas (Bab I)	E-2
Lampiran E-3 : Hasil Uji Similaritas (Bab II)	E-3
Lampiran E-4 : Hasil Uji Similaritas (Bab III)	E-6
Lampiran E-5 : Hasil Uji Similaritas (Bab IV)	E-8
Lampiran E-6 : Hasil Uji Similaritas (Bab V)	E-9
Lampiran E-7 : Hasil Uji Similaritas (Bab VI)	E-11
Lampiran E-7 : Hasil Uji Similaritas (<i>Full</i>)	E-12
LAMPIRAN F :	
<i>Paper</i> dengan Format IEEE	F-1