

Prediksi Pemberian Dana Pengembangan Ke Universitas Menggunakan Model Regresi Logistik

Rico Yoga Pradana^{1*}, Nurmalitasari²

^{1,2}Sistem Informasi

Universitas Duta Bangsa

^{1*}202030191@mhs.udb.ac.id, ² nurmalitasari@udb.ac.id

Abstrak— Perguruan tinggi atau universitas tentu saja ingin mendapatkan suatu penghargaan atau dana pengembangan dari pihak terkait dikarenakan dengan melakukan pemberian dana pengembangan kepada perguruan tinggi dapat meningkatkan sarana dan prasarana universitas dan menghasilkan lulusan mahasiswa dengan ilmu yang kompeten. Penelitian ini digunakan untuk memprediksi penyaluran dana pengembangan ke universitas yang tepat menggunakan model machine learning regresi logistik. Hasil penelitian menunjukkan bahwa keberhasilan siswa dengan menggunakan model regresi skor akurasi model 54%.

Kata kunci— Machine Learning, Regresi Logistik, Sistem Pendukung Keputusan.

Abstract— Colleges or universities certainly want to get an award or development fund from related parties because by providing development funds to universities can improve university facilities and infrastructure and produce graduate students with competent knowledge. This research is used to predict the distribution of development funds to the right university using a logistic regression machine learning model. The outcomes demonstrated that student outcomes may be predicted using the regression model. success with a model accuracy score of 54%.

Keywords— Machine Learning, Logistic Regression, Decision Support Systems.

I. PENDAHULUAN

Ilmu terkait data sains salah satunya adalah prediksi atau peramalan yang bisa digunakan untuk melakukan prediksi diantaranya adalah machine learning, data mining serta lainnya. Beberapa model machine learning digunakan untuk prediksi, salah satunya adalah regresi logistik. Regresi logistik adalah model statistik yang digunakan untuk memprediksi kemungkinan hasil biner, seperti ya atau tidak. Ini adalah jenis pembelajaran yang diawasi, artinya dilatih pada kumpulan data dari data berlabel, di mana setiap titik data memiliki hasil yang diketahui. Selain itu asumsi yang dipenuhi dari regresi logistik ini adalah jumlah data harus besar.

Perguruan tinggi atau universitas tentu saja ingin mendapatkan suatu penghargaan atau dana pengembangan dari pihak terkait dalam hal ini kita contohkan pihak terkait di Indonesia adalah Kementerian Pendidikan Nasional. Dikarenakan dengan melakukan pemberian dana pengembangan kepada perguruan tinggi dapat meningkatkan sarana dan prasarana universitas dan menghasilkan lulusan mahasiswa dengan ilmu yang kompeten.

Ada beberapa kriteria yang mempengaruhi universitas untuk mendapatkan dana pengembangan

seperti faktor Fakultas (CIPDESC), Tingkat Pendidikan (CREDESC), IPEDSCOUNT 1, IPEDSCOUNT 2. Penelitian ini digunakan untuk memprediksi penyaluran dana pengembangan ke universitas yang tepat.

II. METODOLOGI PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini menggunakan metode kualitatif dan menggunakan model regresi logistik.. Variabel dependen yang digunakan dalam penelitian ini adalah indikator universitas yang mendapatkan dana pengembangan selama 1 tahun. Sedangkan variabel independen yang digunakan dalam penelitian ini adalah Fakultas (CIPDESC), Tingkat Pendidikan (CREDESC), IPEDSCOUNT 1, IPEDSCOUNT 2. Data dalam penelitian ini adalah data yang diperoleh dari collegescorecard.ed.gov [1] dengan sample 16484, 233979 universitas yang belum di filter. Model yang digunakan dalam penelitian adalah regresi logistik.

Tabel 1. Variabel *Independen* dan *Dependen* yang digunakan dalam penelitian

Variabel	Nilai Konstruksi Variabel
CIPDESC	1. Accounting, Business & Math
	a. Accounting and Related Services
	b. Mathematics

	c. Mathematics and Statistics, Other		h. Curriculum and Instruction.
	d. Business Administration, Management and Operations		i. Religion/Religious Studies.
	e. Agricultural Business and Management		j. Bilingual, Multilingual, and Multicultural Education.
	a. Finance and Financial Management Services.		k. Rhetoric and Composition/Writing Studies.
	b. International Business.		a. Communication and Media Studies.
2. Science			b. Communication Disorders Sciences and Services.
a. Aerospace, Aeronautical and Astronautical Engineering			c. Missions/Missionary Studies and Missiology.
b. Astronomy and Astrophysics			d. Romance Languages, Literatures, and Linguistics.
c. Biology, General			e. Cosmetology and Related Personal Grooming Services.
d. Cell/Cellular Biology and Anatomical Sciences		6. Technology	
e. Cosmetology and Related Personal Grooming Service		a. Information Science/Studies	
f. Geological and Earth Sciences/Geosciences		b. Radio, Television, and Digital Communication	
g. Geography and Cartography.		c. Computer/Information Technology Administration and Management	
3. Hospitality		d. Energy Systems Technologies/Technicians.	
a. Psychology, General.		e. Drafting/Design Engineering Technologies/Technicians	
b. Physics		f. Electrical and Power Transmission Installers	
c. Health Professions and Related Clinical Sciences, Other		g. Electrical, Electronics and Communications Engineering	
d. Health Services/Allied Health/Health Sciences, General		h. Precision Metal Working.	
e. Health and Medical Administrative Services		7. Art	
f. Health and Physical Education/Fitness		a. Drama/Theatre Arts and Stagecraft	
g. Mental and Social Health Services and Allied Professions		b. Fine and Studio Arts.	
h. Registered Nursing, Nursing Administration, Nursing Research and Clinical Nursing		c. Graphic Communications.	
i. Pharmacology and Toxicology		d. Liberal Arts and Sciences, General Studies and Humanities.	
j. Practical Nursing, Vocational Nursing and Nursing Assistants		e. Visual and Performing Arts, General.	
k. Clinical, Counseling and Applied Psychology		f. Culinary Arts and Related Services.	
l. Allied Health Diagnostic, Intervention, and Treatment Professions.		8. Geology	
m. Allied Health and Medical Assisting Services		a. Area Studies.	
n. Neurobiology and Neurosciences		b. Agriculture, Agriculture Operations, and Related Sciences, Other.	
o. Cognitive Science		c. Natural Resources Conservation and Research.	
4. Society			
a. Anthropology.			
b. Philosophy.			
c. Social Work			
d. Political Science and Government.			
e. Human Resources Management and Services.			
f. History.			
g. Public Administration and Social Service Professions, Other			
h. Homeland Security			
5. Education			
a. Teacher Education and Professional Development, Specific Subject Areas			
b. Teacher Education and Professional Development, Specific Levels and Methods.			
c. Special Education and Teaching.			
d. Education, General.			
e. Educational/Instructional Media Design.			
f. English Language and Literature, General.			
g. Non-Professional General Legal Studies (Undergraduate).			

CREDDESC

1.
 - a. Graduate/Professional Certificate
 - b. First Professional Degree
 - c. Post-baccalaureate Certificate
2.
 - a. Associate's Degree
 - b. Master's Degree
 - c. Doctoral Degree
3.
 - a. Bachelor's Degree
 - b. Diploma

IPEDSCOUNT 1

- 1 = 0 – 39
- 2 = 40 – 79
- 3 = 70 – 109
- 4 = 110 – 149
- 5 = >150

IPEDSCOUNT 2

- 1 = 0 – 39
- 2 = 40 – 79
- 3 = 70 – 109
- 4 = 110 – 149
- 5 = >150

Dalam penelitian ini analisis data menggunakan bahasa pemrograman Python.

Proses dan langkah penelitiannya adalah sebagai berikut:

Step 1 Membersihkan dan memfilter data

Step 2 Pengelompokan variabel

Step 3 Uji chi square variabel.

Step 4 Menentukan persamaan variabel regresi logistik.

Step 5 Melakukan prediksi regresi logistik.

Step 6 Menghitung tingkat prediksi untuk model regresi logistic.

III. HASIL DAN PEMBAHASAN

Step 1. Membersihkan dan memfilter data

Pada tahap ini akan dilakukan filtering dan pembersihan 233979 data dari missing value serta melakukan perhitungan total data yang bisa digunakan.

```
INSTNM      16484
CIPDESC     16484
CREDESC     16484
IPEDSCOUNT1 16484
IPEDSCOUNT2 16484
EARN_COUNT_NWNE_1YR 16484
dtype: int64
```

Gambar 1. Jumlah data variabel yang telah dibersihkan dari missing value

Dari gambar 1 dapat dilihat bahwa dari 233979 data, hanya 16484 data saja yang terbebas dari missing value. Dikarenakan 16484 data ini telah memenuhi syarat maka data ini akan dianalisis ke proses selanjutnya yaitu pengelompokan variabel.

Step 2. Pengelompokan Variabel

Dalam penelitian ini indikator yang digunakan untuk memprediksikan pemberian dana pengembangan berdasarkan data universitas yang pernah mendapatkan dana pengembangan selama 1 tahun. Pada Gambar 2 bisa dilihat bahwa data output yang ditampilkan memiliki dua nilai yang

mana 0 untuk tidak direkomendasikan, dan 1 untuk direkomendasikan.

EARN_COUNT_NWNE_1YR_GR

0.0
0.0
1.0
0.0
1.0
...

Gambar 2. Output EARN_COUNT_NWNE_1YR_GR

Step 3. Uji Chi square variabel

Pengujian hipotesis terhadap perbedaan lebih dari dua proporsi populasi tidak dapat dilakukan dengan distribusi t atau f. Sebaliknya, distribusi Chi-Square dapat digunakan dalam bahasan statistika non parametrik..[2] Pada penelitian kali ini didapatkan variabel yang mendukung untuk bisa digunakan dalam model regresi logistik, seperti yang ada di gambar 3.

Tabel 2. Output dari pengujian variabel yang bisa digunakan

	coef	std err	z	P > z	[0.025	0.975]
const	-3.2640	0.176	-18.556	0.000	-3.609	-2.919
CPD	-0.0266	0.008	-3.277	0.001	-0.043	-0.011
CRD	0.4248	0.030	13.962	0.000	0.365	0.484
IPD1	1.1901	0.179	6.650	0.000	0.839	1.541
IPD2	0.9714	0.175	5.551	0.000	0.628	1.314

Step 4. Menentukan persamaan variabel regresi logistik.

Langkah terpenting dalam membuat prediksi menggunakan regresi logistik adalah menentukan persamaan regresi logistik yang digunakan untuk melakukan prediksi. [3]

Dilihat dari Tabel 3, nilai $P > |z|$ dari semua variabel (Fakultas, Tingkat Pendidikan, IPEDSCOUNT 1, IPEDSCOUNT 2) yang telah ditentukan pada step sebelumnya menunjukkan bahwa $P > |z|$ kurang dari nilai alpha ($\alpha = 5\%$). Nilai $P > |z|$ dari Fakultas (CPD) sebesar $0.001 < 5\%$ maka variabel fakultas memiliki pengaruh dalam pengambilan keputusan dalam pemberian

dana pengembangan ke universitas dengan nilai koefisien -0.0266. Nilai $P > |z|$ dari Tingkat Pendidikan (CRD) sebesar $0.000 < 5\%$ maka variabel fakultas memiliki pengaruh dalam pengambilan keputusan dalam pemberian dana pengembangan ke universitas dengan nilai koefisien 0.4248. Nilai $P > |z|$ dari IPEDSCOUNT 1 (IPD1) sebesar $0.000 < 5\%$ maka variabel fakultas memiliki pengaruh dalam pengambilan keputusan dalam pemberian dana pengembangan ke universitas dengan nilai koefisien 1.1901. Nilai $P > |z|$ dari IPEDSCOUNT 2 (IPD2) sebesar $0.000 < 5\%$ maka variabel fakultas memiliki pengaruh dalam pengambilan keputusan dalam pemberian dana pengembangan ke universitas dengan nilai koefisien 0.9714.

Tabel 3. Output Variabel dalam persamaan

	coef	std err	P > z	Exp(B)
const	-3.2640	0.176	0.000	3.287
CPD	-0.0266	0.008	0.001	2.642
CRD	0.4248	0.030	0.000	1.529
IPD1	1.1901	0.179	0.000	0.974
IPD2	0.9714	0.175	0.000	0.038

Step 5. Melakukan prediksi regresi logistik

Setelah melakukan dan menentukan persamaan variabel pada data untuk model regresi logistik selanjutnya melakukan prediksi dari variabel variabel yang telah ditentukan.

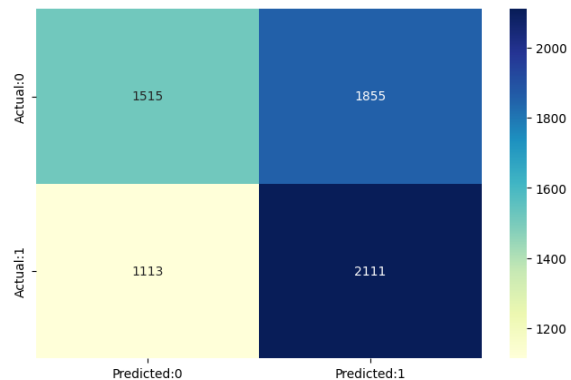
Tabel 4. Hasil prediksi 4% dari populasi yang ada pada data

ID	CPD	CRD	IPD1	IPD2	PR	KMPL
0	3	1	1	1	0.803089	0
1	3	2	1	1	0.796888	0
2	4	2	1	1	0.812262	0
3	3	2	1	1	0.742914	0
4	4	2	1	1	0.746142	0
...	...	...	...	...	...	...
6589	3	3	1	1	0.752602	0
6590	3	2	1	1	0.790626	0
6591	3	3	1	1	0.790688	0
6592	3	3	1	1	0.790626	0
6593	3	2	1	1	0.507796	0

Dilihat pada tabel 2, Nilai probabilitas prediksi yang kurang dari 0.5 direkomendasikan bahwa universitas tersebut layak untuk mendapatkan dana pengembangan, sementara nilai prediksi dengan nilai lebih besar dari 0,5 dianggap kurang direkomendasikan bahwa universitas tersebut layak untuk mendapatkan dana pengembangan.

Step 6. Menghitung tingkat prediksi untuk model regresi logistik.

Setelah melakukan pemodelan regresi logistik maka juga harus memperhitungkan tingkat prediksi model yang telah di buat. (Azizah & Chandra, 2017; Dana & Herlina Samosir dan Made Widiyasa, 2008; Febriani eal., 2017)



Gambar 3. Output dari pengujian variabel yang digunakan

Output confusion matrix yang dihasilkan pada gambar 4, menjelaskan bahwa akurasi model test 40% dari jumlah populasi yang telah dibikin mempunyai TP (Correct result): 1515, FP (Unexpected result): 1855, FN (Missing result): 1113, TN (Correct absence of result): 2111 dan memiliki skor 0.5498938428874734.

IV. KESIMPULAN

Kesimpulan dari penelitian ini adalah bahwa model logistik linear kurang bisa dijadikan patokan untuk dapat membantu dalam pengambilan keputusan pemberian dana pengembangan ke universitas dengan dikarenakan keakurasian skor model regresi logistik 0.5498938428874734.

REFERENSI

- Ari Wibowo. (n.d.). *Uji Chi-Square pada Statistika dan SPSS*.
Azizah, S. M., & Chandra, N. E. (2017). MODEL REGRESI LOGISTIK PADA FAKTOR-FAKTOR YANG MEMPENGARUHI IMUNISASI LENGKAP BALITA. *Jurnal Ilmiah Teknosains*, 3(2).
collegescorecard.ed.gov. (n.d.). *Technical Documentation: College Scorecard Institution-Level Data*.
<https://collegescorecard.ed.gov/assets/CollegeScorecardDataDictionary.xlsx>.
Dana, T., & Herlina Samosir dan Made Widiyasa, D. I. (2008). UPN "Veteran" Yogyakarta. *Seminar Nasional Informatika*.
Febriani, I., Prasetyo, L. B., & Dharmawan, A. H. (2017). Deforestation in Tahura Sekitar Tanjung Province Jambi. *Journal of Natural Resources and Environmental Management*, 7(3), 195–203.
<https://doi.org/10.29244/jpsl.7.3.195-203>
Purwanto, E. (n.d.). *Prediksi Performa Mahasiswa Menggunakan Model Regresi Logistik*.