e-ISBN 978-623-93343-4-5
Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)
Jakarta-Indonesia, 15 September 2021

Memprediksi Kompetensi Kewirausahaan Mahasiswa dengan Metode Klasifikasi *Machine Learning AdaBoost Classifier*

Sabila Rafani Aliandra[1], Aldilla Gardika Pramesta [2],
Matthew Richard Arianto[3], Jamalul Ikhsan[4], Desta Sandya Prasvita[5]
Informatika / Fakultas Ilmu Komputer UPNVJ
Universitas Pembangunan Nasional Veteran Jakarta
Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Kota Depok, Daerah Khusus Ibukota Jakarta 12450
sabilara@upnvj.ac.id [1], aldillagp@upnvj.ac.id [2], matthew@upnvj.ac.id[3], jamaluli@upnvj.ac.id[4],
desta.sandya@upnvj.ac.id[5]

Abstrak. Perkembangan dunia digital di Indonesia begitu cepat, dan terutama perkembangan ekonomi digital yang ikut dalam menyumbang kenaikan ekonomi di Indonesia, salah satunya bidang wirausaha atau seorang entrepreneur. Seorang entrepreneur dapat ditemukan dari berbagai kalangan, dan saat ini yang memiliki potensi besar untuk menjadi seorang entrepreneur itu berasal dari mahasiswa di Universitas. Dan dari potensi tersebut dapat kita ketahui berdasarkan dari kemampuan, kemandirian dan sifat yang lainnya pada mahasiswa tersebut. melakukan suatu penelitian dan membuat sistem pemodelan dengan teknik klasifikasi machine learning AdaBoost Classifier untuk mengetahui potensi atau kapabilitas seorang mahasiswa untuk menjadi seorang entrepreneur yang hebat atau tidak. Pada penelitian ini didapatkan hasil akurasi terhadap evaluasi model pada algoritma Adaboost yaitu sebesar 77.32%.

Kata Kunci: Klasifikasi, Wirausaha, Machine Learning.

1 Pendahuluan

1.1 Latar Belakang.

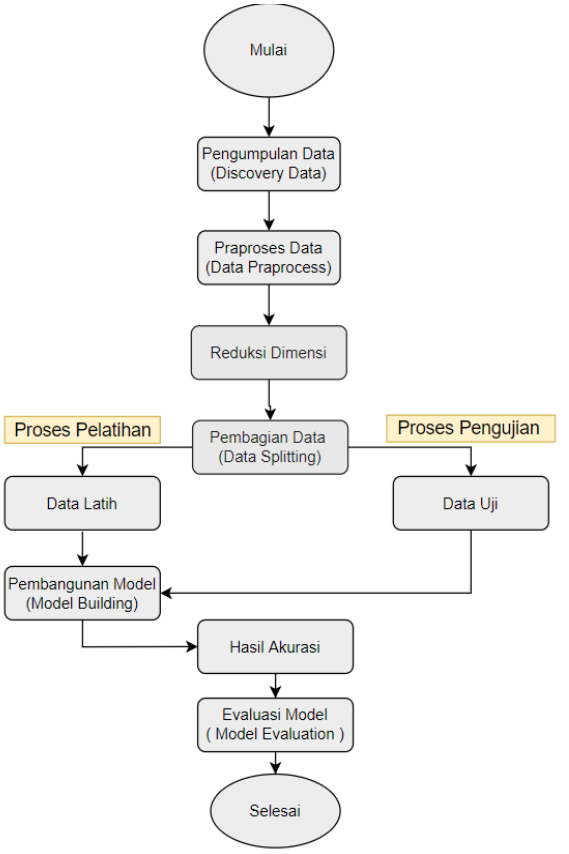
Perkembangan dunia digital di Indonesia begitu cepat, terutama ekonomi digital yang begitu besar menyumbang kemajuan dan kemandirian ekonomi Indonesia. Ekonomi di Indonesia sendiri salah satunya diukur melalui parameter produk domestik bruto (PDB) yang mana kontribusinya berasal dari berbagai aktivitas ekonomi. Peran serta bisnis dan teknologi menunjang ketahanan ekonomi ini, sebagaimana kontribusi terbesar terhadap PDB diberikan oleh sektor usaha mikro. Sektor usaha mikro setiap tahunnya memberikan kontribusi lebih dari 25% terhadap PDB dibandingkan dengan sektor lain yang hanya mampu memberikan kontribusi di bawah 20% [1]. UMKM di Indonesia telah menjadi bagian penting dari sistem perekonomian di Indonesia. Hal ini dikarenakan UMKM merupakan unit-unit usaha yang lebih banyak jumlahnya dibandingkan usaha industri berskala besar dan memiliki keunggulan dalam menyerap tenaga kerja lebih banyak dan juga mampu mempercepat proses pemerataan sebagai bagian dari pembangunan [2].

Tentunya para aktor perekonomian menjadi salah satu kunci kesuksesan kemandirian ekonomi. Hal ini turut melibatkan peran serta dan dukungan perguruan tinggi yang cermat dalam melihat potensi serta mendidik dan menghasilkan mahasiswa yang unggul dalam kompetensi kewirausahaan. Mahasiswa yang mampu menjadi seorang pebisnis atau entrepreneur adalah mahasiswa yang berpeluang menjadi aktor kunci di masa depan perekonomian Indonesia, terutama dalam mengelola bonus demografi, yakni masa emas Indonesia di 2045 [3].

Berdasarkan dari latar belakang tersebut, dilakukan penelitian untuk mengetahui potensi kemampuan wirausaha seorang mahasiswa dengan merujuk dari penelitian sebelumnya yang didapatkan yaitu dari Predicting and Improving Entrepreneurial Competency in University Students using Machine Learning Algorithms (2020), penelitian rujukan ini memiliki hasil tingkat akurasi sebesar 59,18% menggunakan Support Vector Machine (SVM) dalam melakukan prediksi terhadap data [4]. Selain itu, penelitian rujukan lainnya berasal dari Mental Stress Detection in University Students using Machine Learning Algorithms yang mana Peneliti memprediksi perilaku wirausaha individu dengan menerapkan teori perilaku terencana dengan menerapkan 10-Fold Cross-Validation dan akurasi tertinggi yang dicatat oleh Support Vector Machine (85,71%) [5], sehingga akan dilakukan optimasi dan perbaikan dengan mengganti teknik klasifikasi dengan menggunakan klasifikasi AdaBoost untuk dapat

meningkatkan tingkat akurasi dan hasil yang lebih baik dari penelitian sebelumnya.

2 Metodologi Penelitian



Gambar. 1. Flowchart Metodologi Penelitian

2.1 Pengumpulan Data *(Discovery Data)*

Dalam proses pertama pengumpulan data yaitu data yang digunakan dalam kegiatan ini didapatkan dari sebuah platform dataset kaggle dengan data berjudul "entrepreneurial-competency-in-university-students". Dan data yang didapatkan berupa bentuk csv. Kemudian data ini nantinya akan diolah dan diproses menggunakan algoritma klasifikasi AdaBoost untuk mendapatkan sebuah model dari data tersebut. Kemudian setelah data didapatkan data akan dilakukan proses pembersihan atau preprocessing sebelum data digunakan untuk proses berikutnya yaitu pembangunan model dengan data latih dan data uji. Adapun informasi variabel pada dataset berikut ini.

Tabel 1. Infromasi Variabel pada Dataset

Column	Keterangan
EducationSector	Latar belakang pendidikan.
IndividualProject	Mahasiswa memiliki project pribadi.
Age	Umur Mahasiswa.
Gender	Jenis kelamin Mahasiswa.
City	Tempat tinggal Mahasiswa.
Influenced	Jika Mahasiswa dipengaruhi oleh seseorang.

Perseverance	Peringkat seorang siswa berdasarkan ketekunan.
DesireToTakeInitiative	Peringkat siswa berdasarkan keinginan untuk mengambil inisiatif.
Competitiveness	Peringkat kompetitif
SelfReliance	Peringkat kemandirian
StrongNeedToAchieve	Kebutuhan yang kuat untuk mencapai peringkat tujuan
SelfConfidence	Peringkat kepercayaan diri
GoodPhysicalHealth	Peringkat kesehatan fisik yang baik
MentalDisorder	Jika ada gangguan jiwa
KeyTraits	Ciri-ciri utama siswa
ReasonsForLacky	Alasan kurangnya budaya kewirausahaan
	Kelas (0 atau 1)

2.2 Praproses Data *(Data Praprocess)*

Data yang digunakan dalam proses klasifikasi ini data entrepreneurial-competency-in-university-students berupa file dengan tipe csv. Dan didalam file csv ini terdapat data yang terdiri dari sebanyak 219 baris data dan 17 kolom serta 2 kelas yaitu untuk klasifikasi great of entrepreneur yang tergolong ke dalam seorang entrepreneur yang baik atau tidak. Dalam penelitian ini setelah mendapatkan data , kemudian dilakukan konversi data menjadi sebuah data frame yang siap untuk diolah menggunakan bahasa pemrograman python. Kemudian melakukan analisis data yang terdapat pada dataset tersebut, dan menemukan bahwa data yang ada, tidak memiliki standar yang sama, dalam dataset terdapat 8 fitur data kategorik dan 9 fitur data numerik sehingga dilakukan transformasi terlebih dahulu menggunakan label encoder dan data dummies. Normalisasi data juga dilakukan pada dataset menggunakan StandardScaler.

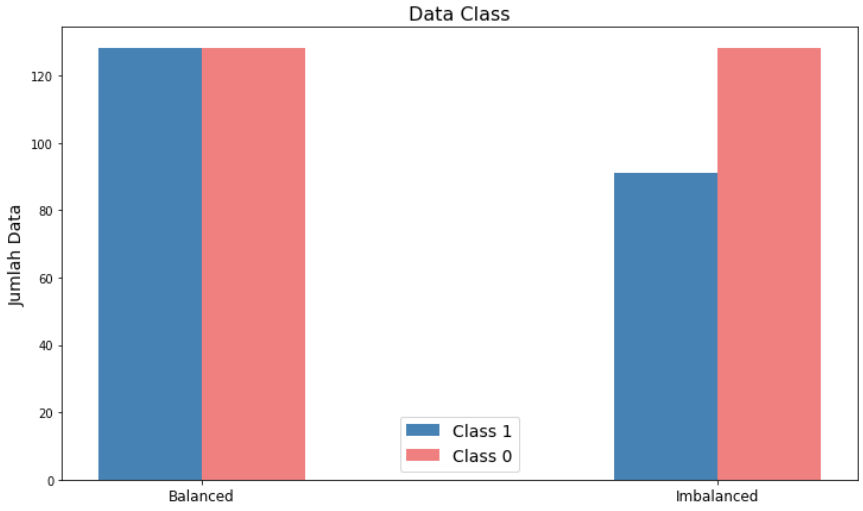
	EducationSector	IndividualProject	Age	Gender	City	Influenced	Perseverance	DesireToTakeInitiative	Competitiveness	SelfReliance	StrongNeedToAchieve	SelfConfidence	GoodPhysi
0	Engineering Sciences	No	19	Male	Yes	No	2	2	3	3	2	2	
1	Engineering Sciences	Yes	22	Male	No	Yes	3	3	3	4	4	3	
2	Engineering Sciences	No	18	Male	Yes	No	3	4	3	3	3	4	
3	Engineering Sciences	Yes	20	Male	Yes	Yes	3	3	3	3	4	3	
4	Engineering Sciences	Yes	19	Male	Yes	Yes	2	3	3	3	4	3	
5	Engineering Sciences	No	19	Male	Yes	Yes	3	3	3	3	3	3	
6	Engineering Sciences	Yes	19	Male	Yes	Yes	3	2	3	3	4	1	
7	Engineering Sciences	No	20	Male	Yes	Yes	4	2	4	4	5	3	
8	Others	Yes	20	Male	Yes	Yes	2	3	3	1	2	2	
9	Engineering Sciences	Yes	17	Male	Yes	Yes	2	3	2	1	4	3	

Gambar. 2. Sebelum dilakukan Label Encoder

	EducationSector	Age	Perseverance	DesireToTakeInitiative	Competitiveness	SelfReliance	StrongNeedToAchieve	SelfConfidence	GoodPhysicalHealth	KeyTraits	y	IndividualProject_No
0	2	19	2	2	3	3	2	2	3	0	1	1
1	2	22	3	3	3	4	4	3	4	3	0	0
2	2	18	3	4	3	3	3	4	4	0	0	1
3	2	20	3	3	3	3	4	3	3	2	0	0
4	2	19	2	3	3	3	4	3	2	3	1	0
5	2	19	3	3	3	3	3	3	3	1	1	1
6	2	19	3	2	3	3	4	1	1	4	1	0
7	2	20	4	2	4	4	5	3	4	4	0	1
8	7	20	2	3	3	1	2	2	2	0	0	0
9	2	17	2	3	2	1	4	3	3	4	1	0

Gambar. 3. Setelah dilakukan Label Encoder

Selain itu, dilakukan proses *exploratory data analysis* (EDA) dan ditemukan satu fitur dengan jumlah missing value cukup besar, maka fitur tersebut dihapus. Kemudian didapatkan kondisi jumlah data kelas tidak seimbang (*imbalanced*). Jumlah data dengan kelas 0 adalah 128, sedangkan kelas 1 berjumlah 91. Oleh karena itu dilakukan *oversampling* untuk menyeimbangkan datanya dengan metode random oversampling/resample agar data kelas yang minoritas sama dengan kelas mayoritas, sehingga didapatkan jumlah data dengan kelas 1 sama dengan kelas 0, yaitu 128. Kini jumlah total *record* data adalah 256.



Gambar 4. Hasil Kelas Imbalanced Dan Balanced

2.3 Reduksi Dimensi

Analisis komponen utama (PCA) adalah teknik yang digunakan untuk menyederhanakan suatu data, dengan cara mentransformasi linier sehingga terbentuk sistem koordinat baru dengan variansi maksimum. PCA dapat digunakan untuk mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara signifikan. Metode ini mengubah dari sebagian besar variabel asli yang saling berkorelasi menjadi satu himpunan variabel baru yang lebih kecil dan saling bebas (tidak berkorelasi lagi). Dalam penelitian ini terdapat 15 buah fitur yang digunakan untuk memprediksi kompetensi kewirausahaan mahasiswa. Jumlah ini tentunya masih cukup banyak dan sulit divisualisasikan. Oleh sebab itu, data dilakukan reduksi dimensi dengan metode Principal Component Analysis hingga diperoleh 2 fitur saja.

2.4 Pembagian Data *(Data Splitting)*

Kemudian data yang sudah siap akan dibagi menjadi dua bagian yaitu data untuk pelatihan dan data untuk pengujian. Dimana masing-masing data memiliki proporsi sebanyak 80% untuk data latih dan 20% untuk data uji. Pada penelitian ini dipilih proporsi data tersebut agar data yang diproses oleh algoritma AdaBoost tersebut dapat

menghasilkan sebuah model yang baik. Sebelum data dibagi menjadi data train dan data test dilakukan pengacakan data atau random sample/shuffle sampel agar data yang diproses bisa lebih baik dan proporsional.

2.5 Pembangunan Model *(Model Building)*

Tabel 2. Algoritma Adaboost

Algoritma 1: Adaboost

1. Terdapat m data latih :

$(x_1, y_1), \dots, (x_m, y_m); x_i \in x, y_i \in \{-1, +1\}$

2. Inisialisasi bobot awal

$$D_1(i) \leftarrow \frac{1}{m}; i, \dots, m$$

3. Untuk t=1,…,T lakukan langkah 4 sampai 8:

4. Bangun classifier menggunakan weak classifier

terhadap distribuasi data D_t

5. Hitung ε_t :

$$\varepsilon_t \leftarrow \sum_{i:M^t(x^i)\neq y^i} D_t(i)$$

Jika $\varepsilon_t > 0.5$ maka berhenti

6. Hitung nilai α_t

$$\alpha_t \leftarrow \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$$

7. Update nilai $D_{t+1}(i)$

$$D^{t+1}(i) \leftarrow D^t(i).\, e^{-\alpha^t y^t M^t(x^i)}$$

8. Normalisasi D_{t+1}

9. Classifier kuat

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t M_t(x)\right)$$

Data yang digunakan dalam proses pembangunan model ini yaitu sebanyak 80% data latih dari total data yang terdapat pada dataset atau sekitar 205 dari 256 total data. Kemudian data latih ini akan dilakukan pelatihan secara terus-menerus oleh algoritma klasifikasi AdaBoost sisanya data akan digunakan untuk proses data pengujian atau data testing.

2.6 Evaluasi Model *(Model Evaluation)*

Untuk mengevaluasi atau validasi dari model yang sudah dihasilkan dari algoritma AdaBoost ini. Pada penelitian ini digunakan evaluasi model dengan K-fold Cross Validation. Dari akurasi yang sudah dihasilkan juga dapat ditingkatkan akurasi nya dengan mengubah atau menambahkan parameter dalam evaluasi model tersebut.

2.6.1 K-fold Cross Validation

K-fold adalah salah satu metode Cross Validation yang populer dengan melipat data sebanyak K dan mengulangi (iterasi) eksperimennya sebanyak K juga. Pada pengujian ini data yang akan digunakan adalah 256 data termasuk data uji yang nantinya akan dibagi menjadi 10 bagian atau k=10 sehingga data yang diperoleh adalah 256 data dibagi menjadi 10 lipatan. Selain itu akan ditentukan mana yang termasuk data training dan mana yang termasuk data testing dengan perbandingan 9:1. Pengujian menggunakan data yang sudah dipartisi akan diulang sebanyak 10 kali (K=10) dengan posisi data testing berbeda di setiap iterasinya. Misalkan iterasi pertama data tes pada posisi awal, iterasi kedua data testing di posisi kedua begitu seterusnya.

Tabel 3. Ilustrasi 10-fold Cross Validation

Iterasi 1/10	**Test Set**								
Iterasi 2/10		**Test Set**							
Iterasi 3/10			**Test Set**						
Iterasi 4/10				**Test Set**					
Iterasi 5/10					**Test Set**				
Iterasi 6/10						**Test Set**			
Iterasi 7/10							**Test Set**		
Iterasi 8/10								**Test Set**	
Iterasi 9/10									**Test Set**
Iterasi 10/10									**Test Set**

3 Hasil Penelitian

Hasil dari penelitian rujukan sebelumnya yaitu pada paper penelitian Predicting and Improving Entrepreneurial Competency in University Students using Machine Learning Algorithms memiliki hasil tingkat akurasi dibawah 60% dari total 6 algoritma klasifikasi yang telah dilakukan terhadap data seperti yang terlihat pada gambar tabel dibawah ini. Dari total 6 algoritma klasifikasi tersebut, yang memiliki hasil yang paling tinggi yaitu diperoleh dari algoritma Support Vector Machine sebesar 59.18%

Tabel 4. Hasil Akurasi Penelitian Rujukan

No	Algoritma	Akurasi
1	Random Forest (500 Trees)	40.81%
2	K-Nearest Neighbors (k = 5)	53.06%

3	Support Vector Machine	59.18%
4	Logistic Reggression	57.14%
5	Naïve Bayes	48.97%
6	Decision Tree	57.14%

Hasil penelitian ini yang dilakukan proses pengolahan terhadap dataset entrepreneur tersebut menggunakan algoritma klasifikasi AdaBoost memperoleh tingkat akurasi sebesar 75%.

Tabel 5. Akurasi Model

No	Algoritma	Akurasi
1	Adaboost Classifier	75%

Kemudian, dari hasil tingkat akurasi model tersebut dilakukan validasi model untuk mengetahui performa rata-rata model yang dihasilkan oleh algoritma klasifikasi Adaboost tersebut menggunakan K-Fold Cross Validation dengan K yaitu 10. Rata-rata akurasi dari semua nilai K yang dihasilkan sebesar 77.32%.

Tabel 6. K-Fold Cross Validation (K = 10)

K	Akurasi
1	84.61%
2	76.92%
3	76.92%
4	84.61%
5	80.76%
6	65.38%
7	72%
8	68%
9	84%
10	80%

4 Kesimpulan

Dari kegiatan penelitian ini dan berdasarkan dari hasil di atas antara lain , keberhasilan melakukan proses peningkatan akurasi terhadap data entrepreneur yang digunakan pada paper penelitian sebelumnya dimana hasil tingkat akurasi pada paper penelitian sebelumnya hanya mampu mencapai angka paling tinggi yaitu sebesar 59.18%, dan hasil akurasi yang didapatkan pada penelitian ini yaitu sebesar 77.32%.

Tahapan pada saat pra proses data sangat menentukan hasil dari akurasi klasifikasi, kemudian melakukan proses standarisasi untuk menyetarakan semua data dengan jangkauan angka yang sama, reduksi dimensi untuk meminimalkan kompleksitas, dan terakhir dilakukan proses imbalance data dengan random oversampling terhadap data target untuk menghindari hasil akurasi yang buruk.

5 Saran

Adapun penelitian yang telah dilakukan ini masih memiliki kekurangan untuk hasil akurasi yang didapatkan , sehingga saran untuk penelitian berikutnya dapat melakukan pengembangan untuk pengolahan data pada saat praproses data menggunakan metode praproses yang lebih baik dan tentunya menggunakan algoritma yang dapat melakukan prediksi dengan hasil akurasi yang jauh lebih baik lagi.

Referensi

[1] Santosa, T., & Budi, Y. R. (2020). Analisa Perkembangan UMKM di Indonesia Pada Tahun 2017-2019. Develop: Jurnal Ekonomi Pembangunan, 1(2), 57-64.
[2] Suci, Y. R. (2017). Perkembangan UMKM (Usaha mikro kecil dan menengah) di Indonesia. Jurnal Ilmiah Cano Ekonomos, 6(1), 51-58.
[3] Abi, A. R. (2017). Paradigma Membangun Generasi Emas Indonesia Tahun 2045. Jurnal Ilmiah Pendidikan Pancasila dan Kewarganegaraan, 2(2), 85-90.
[4] Sharma, U., & Manchanda, N. (2020, January). Predicting and Improving Entrepreneurial Competency in University Students using Machine Learning Algorithms. In *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 305-309). IEEE.
[5] Ahuja, R., & Banga, A. (2019). Mental stress detection in university students using machine learning algorithms. *Procedia Computer Science*, *152*, 349-353.
[6] Rokach, L. Ensemble-based classifiers. *Artif Intell Rev* 33, 1–39 (2010).
[7] Prasvita, D. S., Komp, S., & Kom, M. METODE ADABOOST PADA SKEMA PEMODELAN HYBRID UNTUK KLASIFIKASI PENYAKIT LIVER.
[8] https://www.kaggle.com/namanmanchanda/entrepreneurial-competency-in-university-students. (Diakses 22 Juni 2021)