

Prediksi Kesuksesan *Startup* Menggunakan *Deep Neural Network*

Wulan Sri Lestari^{1*}, Apriyanto Halim²

^{1,2}Universitas Mikroskil, Jl. Thamrin No. 112, 124, 140, Telp. (061) 4567789, Fax. (061) 4567789

^{1,2}Fakultas Informatika, Program Studi Teknologi Informasi, Universitas Mikroskil, Medan

¹wulan.lestari@mikroskil.ac.id, ²apriyanto.halim@mikroskil.ac.id

Abstrak

Kesuksesan *startup* memiliki peran penting dalam pertumbuhan ekonomi dengan ide bisnis yang baru, inovatif dan dapat menciptakan lapangan pekerjaan. Pertumbuhan *startup* yang eksponensial dalam beberapa tahun terakhir membuat ketidakpastian dan tingkat kegagalan yang tinggi. Sehingga penting bagi investor untuk dapat memprediksi kesuksesan *startup* dalam menemukan perusahaan yang memiliki potensi kesuksesan yang lebih besar untuk didanai. Untuk membantu para investor, maka tujuan dari penelitian ini adalah membangun model untuk memprediksi apakah *startup* yang sedang beroperasi akan sukses atau gagal menggunakan *Deep Neural Network* (DNN). DNN mengkombinasikan keunggulan *deep learning* dan *neural network* untuk memecahkan masalah *nonlinear*. Proses yang dilakukan adalah menggunakan *dataset startup success prediction*, kemudian *dataset* tersebut di *pre-processing* untuk pengecekan *missing value*, *data duplicate*, *data outlier* serta penentuan atribut berdasarkan korelasi antar variabel. Kemudian, *dataset* yang sudah di *pre-processing* dibagi menjadi data *training* dan data *testing*. Selanjutnya ditentukan parameter DNN apa saja yang akan digunakan untuk membangun model prediksi menggunakan data *training* agar tidak mengalami *overfitting* ataupun *underfitting*. Model yang dibangun kemudian diuji menggunakan data *testing*. Hasil pengujian menunjukkan bahwa model prediksi yang diusulkan tidak *overfitting* ataupun *underfitting* dan memiliki nilai akurasi sebesar 83,93% dan nilai presisi sebesar 86,51% dalam memprediksi kesuksesan *startup*.

Kata kunci: Prediksi Kesuksesan, *Startup*, *Deep Neural Networks*

Abstract

Startup success has an important role in economic growth with new, innovative business ideas that can create jobs. The exponential growth of startups in recent years has created uncertainty and high failure rates. Investors need to predict startup success by finding companies with greater potential for success to fund. This research aims to help investors build a model to predict whether startups currently operating will be successful or fail using Deep Neural Networks. DNN combines the advantages of deep learning and neural networks to solve nonlinear problems. The process that is carried out is using the startup success prediction dataset, then the dataset is preprocessed to check for missing values, duplicate data, outlier data and determine attributes based on correlations between variables. Then, the preprocessed dataset is divided into training data and testing data. Next, it is determined which DNN parameters will be used to build a predictive model using training data so that it does not experience overfitting or underfitting. The built model is then tested using data testing. The test results show that the proposed prediction model is neither overfitting nor underfitting and has an accuracy value of 83.93% and a precision value of 86.51% in predicting startup success.

Keywords: Success Prediction, *Startup*, *Deep Neural Networks*

1. PENDAHULUAN

Startup merupakan sebuah perusahaan rintisan yang terdiri dari sekumpulan individu untuk mengembangkan produk atau jasa dengan memanfaatkan teknologi informasi [1]. Kesuksesan *startup*

memiliki pengaruh yang besar dalam mendorong perkembangan ekonomi [2]. Namun, banyaknya jumlah *startup* yang ada tidak semuanya mengalami kesuksesan. Secara umum, 90 dari 100 *startup* gagal mendapatkan pendanaan yang diharapkan [3]. Ada banyak faktor yang menyebabkan kegagalan sebuah *startup* seperti manajemen yang buruk, pendanaan yang kurang, kerja sama tim dan sebagainya [3]. Kegagalan *startup* dapat menimbulkan kerugian bagi semua orang yang berkontribusi termasuk salah satunya adalah *venture capital* sebagai investor. Oleh karena itu, kemampuan *venture capital* dalam memprediksi kesuksesan *startup* merupakan bagian penting bagi investor untuk menentukan apakah *startup* tersebut layak didanai atau tidak, sehingga dibutuhkan model prediksi kesuksesan *startup*.

Terdapat beberapa penelitian yang sudah dilakukan terkait prediksi kesuksesan *startup*. Tahun 2021, Permana *et al.*, melakukan penelitian prediksi kesuksesan *startup* menggunakan *Decision Tree*, *kNN*, dan *Naïve Bayes* dengan nilai akurasi sebesar 79,29%, 66,69% dan 64,21% [4]. Pada penelitian [4] jumlah atribut yang digunakan untuk membangun model prediksi adalah 19 atribut dari total 49 atribut yang ada pada *dataset*. Kemudian dilakukan proses *pre-processing* data untuk mengatasi masalah *missing value* dengan mengisi setiap nilai atribut yang kosong dengan nilai rata-rata atribut. Goldenia *et al.*, juga melakukan penelitian terkait prediksi kesuksesan *startup* menggunakan *Support Vector Machine* (SVM) dengan nilai akurasi sebesar 79,1% [5]. Proses yang dilakukan oleh [5] adalah melakukan *pre-processing* data untuk mengatasi *missing value*, data yang duplikasi dan *outlier* serta menggunakan 32 atribut dari total 49 atribut yang ada pada *dataset* dan menambahkan 4 atribut tambahan yaitu *has_RoundABCD*, *has_Investor*, *has_Seed*, dan *age_startup_year*. Kedua penelitian tersebut [4] dan [5] menggunakan *dataset* yang sama yang akan digunakan pada penelitian ini, sehingga hasil akurasi yang didapatkan dari kedua penelitian sebelumnya [4] dan [5] masih dapat ditingkatkan. Untuk meningkatkan hasil akurasi prediksi kesuksesan *startup* maka pada penelitian ini digunakan *Deep Neural Network* (DNN).

DNN merupakan teknologi pembelajaran mesin yang didasarkan pada konsep *Artificial Neural Network*. DNN mengkombinasikan keunggulan *deep learning* dan *neural network* untuk memecahkan masalah *nonlinear* dengan lebih baik dibandingkan dengan algoritma mesin *learning* konvensional [6]. Selain itu, [7] melakukan penelitian terkait prediksi stroke menggunakan DNN dengan hasil akurasi 96% dan [8] menggunakan DNN untuk prediksi kecacatan material dengan nilai akurasi hasil *testing* sebesar 93%. Proses yang dilakukan pada penelitian ini adalah melakukan *pre-processing* data sebelum data tersebut digunakan untuk membangun model prediksi menggunakan DNN. Pada tahap *pre-processing* dilakukan proses pengecekan *missing value*, duplikasi data, pengecekan *outlier* dan penentuan atribut yang akan digunakan dalam pembuatan model prediksi. *Dataset* yang digunakan pada penelitian ini sama dengan penelitian Goldenia *et al.* Hal ini dilakukan karena fokus/kontribusi pada penelitian ini adalah kebaruan metode untuk menghasilkan model prediksi kesuksesan *startup* yang tidak *overfitting* ataupun *underfitting* dengan akurasi yang baik.

2. TINJAUAN PUSTAKA

2.1 *Startup*

Startup merupakan sebuah perusahaan rintisan yang terdiri dari sekumpulan individu untuk mengembangkan produk atau jasa dengan memanfaatkan teknologi informasi [1]. *Startup* dibangun dalam kondisi yang tidak pasti sehingga memiliki risiko kegagalan yang tinggi karena produk atau jasa yang ditawarkan tidak sesuai dengan kebutuhan pengguna atau kalah bersaing dengan perusahaan sejenis [5].

Untuk memulai sebuah *startup* terdapat beberapa tahapan yang umum dilakukan sebagai berikut [9]:

a. *Idetation*

Tahapan *idetation* merupakan tahap yang fokus untuk mencari dan menemukan ide yang layak dijadikan sebuah bisnis dan akan berkembang dimasa depan. Tahap *idetation* dimulai dari mencari masalah, mencari solusi dari masalah, mencari dan membandingkan dengan solusi yang sudah ada dan menentukan *timing* yang tepat untuk eksekusi ide.

b. *Product Development*

Setelah ide didapatkan maka selanjutnya adalah pengembangan ide menjadi sebuah produk. Masalah utama *startup* dalam mengembangkan produk adalah keterbatasan sumber daya. Untuk mengatasi keterbatasan tersebut, ada beberapa strategi pengembangan produk yang dapat dilakukan seperti membuat produk sendiri, mencari mitra (*co-founder*), *outsource* dan membayar karyawan. Saat pengembangan produk pastikan *startup* minimal sampai pada versi MVP (*Minimum Viable Product*) agar layak digunakan di tahap awal.

c. Model Bisnis

Sejalan dengan pengembangan produk, kita perlu memikirkan model bisnis seperti apa yang akan digunakan dalam *startup*. Model bisnis tidak hanya terkait dengan bagaimana mendapatkan uang dari pelanggan, namun juga berkaitan dengan pihak-pihak yang terlibat untuk mengantarkan produk sampai kepada pelanggan. Salah satu *tools* yang dapat digunakan untuk mendiskusikan model bisnis adalah *business canvas*. Berikut ini adalah beberapa model pembayaran yang dapat dilakukan oleh pelanggan kepada *startup*.

1. *On time*: Pelanggan membayar cukup satu kali dan produk menjadi milik pelanggan.
2. *Subscription*: Pelanggan membayar sesuai dengan waktu yang ditentukan apakah per bulan atau per tahun, meskipun layanan digunakan atau tidak.
3. *Pay by advertisement*: Layanan dibayar oleh iklan dan pelanggan tidak membayar apa pun penggunaan produk.
4. *Pay as you go*: Pelanggan hanya membayar jika saat menggunakan produk saja jika tidak maka tidak perlu membayar apa pun.
5. *Barter*: Layanan ditukar dengan sesuatu seperti kuota data atau aktivitas lainnya.
6. *Freemium*: Pelanggan bebas menggunakan produk tanpa membayar apa pun.

Tidak ada satu pun model bisnis yang cocok untuk diterapkan ke semua produk. Anda perlu teliti dalam menentukan mana model bisnis yang paling sesuai dengan produk Anda.

d. *Marketing*

Tahap selanjutnya adalah memperkenalkan produk kepada calon pelanggan. Pada saat melakukan *marketing* maka kita dapat memanfaatkan teknologi *digital marketing*, *search engine optimization* dan *pirate metric*. *Pirate metric* terdiri dari *Acquisition*, *Activation*, *Retention*, *Referral*, dan *Revenue*.

e. Pendanaan

Salah satu aspek penting dalam pengembangan *startup* adalah pendanaan. Ada banyak sumber pendanaan yang bisa dilakukan seperti mendanai sendiri, mencari *angel investor*, mencari *institutionalized investor* dan evaluasi (berapa nilai *startup* anda).

f. *Pitching*

Pitching tidak harus selalu berorientasi kepada finansial, seperti meminta pendanaan kepada calon investor tersebut, tetapi juga dapat digunakan untuk melempar ide kepada calon pengguna, *co-founder*, atau pihak-pihak lain. Pada dasarnya *pitching* adalah melakukan presentasi. Ada beberapa kesalahan yang sering dilakukan pada saat *pitching* seperti tidak memahami siapa pendengarnya, tidak menjelaskan masalah yang ingin diselesaikan, tidak memahami apa yang sudah dilakukan oleh orang lain dalam menyelesaikan masalah sejenis dan solusi yang ditawarkan tidak unik.

2.2 Kesuksesan *Startup*

Startup dikategorikan berhasil ketika mampu menemukan strategi yang tepat untuk pertumbuhan *startup* tersebut [10]. Sebuah *startup* dikatakan sukses apabila mendapatkan pendanaan dalam jumlah besar melalui proses *merger* dan akuisisi atau *initial public offering* (IPO). Sedangkan *startup* dikatakan gagal apabila tidak mampu untuk bertumbuh dan menghasilkan keuntungan [10]. Produk atau jasa yang dibangun sesuai dengan permintaan pasar dengan mengutamakan efisiensi dan efektivitas merupakan salah satu kunci kesuksesan sebuah *startup*. Selain itu, tujuan dan sasaran perusahaan juga menjadi tolak ukur kesuksesan sebuah *startup*. Namun, untuk *startup* berbasis teknologi, kesuksesan usaha adalah ketika *startup* tersebut mampu mencapai *product-market-fit* atau produk yang mereka tawarkan sesuai dengan permintaan pasar [10]. Salah satu cara yang dapat dilakukan untuk mendukung kesuksesan *startup* adalah dengan mengikuti program inkubasi *startup* dengan menerapkan konsep *knowledge sharing* untuk meningkatkan inovasi [11]. Penerapan *knowledge sharing* di perusahaan dapat memberikan

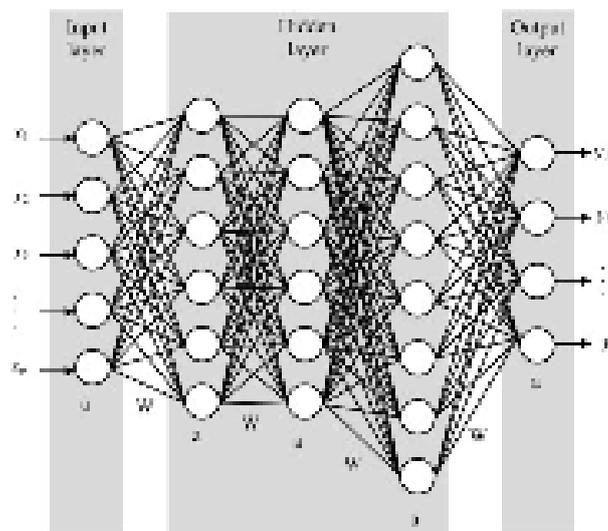
banyak keuntungan seperti meningkatkan kemampuan untuk melihat kesempatan lebih cepat sehingga inovasi dapat dilakukan dengan lebih cepat tanpa menghabiskan banyak biaya operasional [12]. Gambar 1 merupakan contoh-contoh *startup* yang sudah sukses menjadi *unicorn* di Asia Tenggara.



Gambar 1. *Startup Unicorn* Asia Tenggara [12]

2.3 Deep Neural Networks

Deep Neural Networks (DNN) merupakan teknologi pembelajaran mesin yang didasarkan pada konsep *Artificial Neural Network*. DNN mengkombinasikan keunggulan *deep learning* dan *neural network* untuk memecahkan masalah *nonlinear* dengan lebih baik dibandingkan dengan algoritma mesin *learning* konvensional [6]. DNN dibangun dengan banyak lapisan dan setiap lapisan terhubung ke lapisan berikutnya, kecuali untuk lapisan *input* dan *output*. Gambar 2 merupakan struktur DNN.

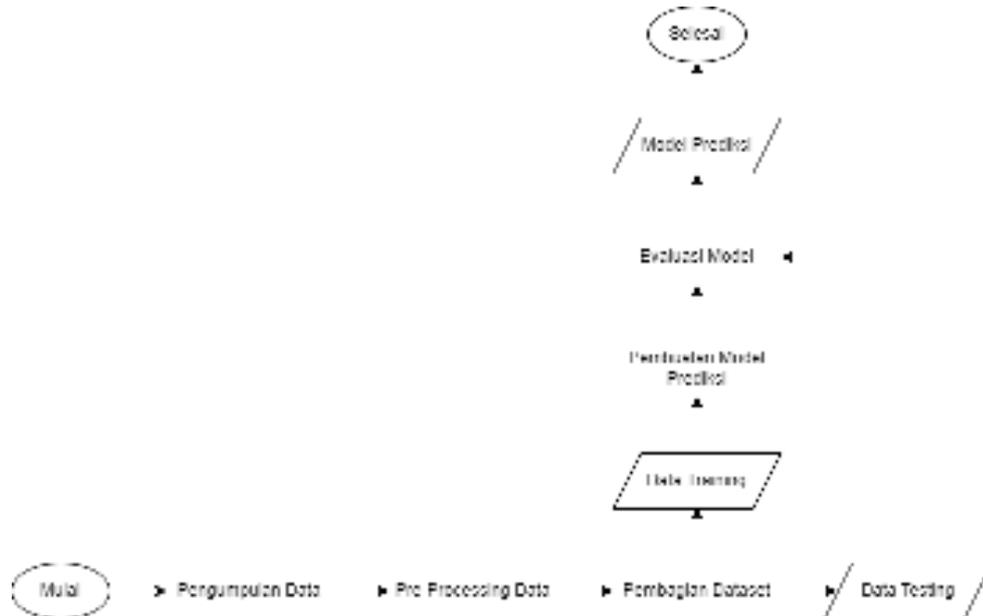


Gambar 2. Struktur DNN [13]

Dalam DNN, setiap *neuron* pada setiap lapisan terhubung ke masing-masing *neuron* pada lapisan berikutnya, seperti yang ditunjukkan pada Gambar 2. Jumlah *neuron* pada lapisan *input* harus sama dengan dimensi fitur. Jumlah *neuron* pada *layer output* tergantung pada berapa banyak label berbeda yang dimiliki *dataset*. Misalnya, jika hanya ada satu *neuron* pada *layer output*, DNN dapat digunakan sebagai klasifikasi biner, karena 0 mewakili *false*, dan 1 mewakili *true*.

3. METODE PENELITIAN

Tahapan penelitian yang dilakukan pada dapat dilihat pada Gambar 3.



Gambar 3. Tahapan Model Prediksi

3.1 Pengumpulan Data

Pada tahap ini dilakukan proses pengumpulan data yang digunakan dari *website kaggle* [14]. *Dataset Startup Success Prediction* merupakan *dataset* dari kumpulan perusahaan *startup* yang berada di negara Amerika Serikat. *Dataset* ini memiliki jumlah 923 *record* data dan terdapat 597 *startup* yang berstatus *acquired* dan 326 *startup* berstatus *closed*. *Dataset* terdiri dari 49 variabel, dengan 48 variabel sebagai atribut dan 1 variabel (status) sebagai kelas yang jika status *startup* adalah *acquired* maka *startup* dinyatakan sukses dan jika *closed* maka *startup* dinyatakan gagal. Tabel 1 merupakan *dataset* yang belum dikelola.

Tabel 1. *Dataset* Yang Belum Dikelola

No	Unna med: 0	state_code	latitude	longitude	zip_code	id	city	...	Status
1	1005	CA	42.35888	-71.05682	92101	c:6669	San Diego	...	acquired
2	204	CA	37.238916	-121.973718	95032	c:16283	Los Gatos	...	acquired
3	1001	CA	32.901049	-117.192656	92121	c:65620	San Diego	...	acquired
4	738	CA	37.320309	-122.05004	95014	c:42668	Cupertino	...	acquired
5	1002	CA	37.779281	-122.419236	94105	c:65806	San Francisco	...	Closed

6	379	CA	37.406914	-122.09037	94043	c:22898	Mount ain View	...	Closed
7	195	CA	37.391558 9	- 122.070264 3	94041	c:16191	Mount ain View	...	acquired
8	875	CA	38.057107	-122.513742	94901	c:5192	San Rafael	...	acquired
9	16	MA	42.712207	-73.203599	1267	c:1043	Willia mstow n	...	acquired
10	846	CA	37.427235	-122.145783	94306	c:498	Palo Alto	...	acquired
...
923	462	CA	37.386778	-121.966277	95054	c:26702	Santa Clara	...	acquired

3.2 Pre-processing Data

Tahap *pre-processing* data merupakan salah satu tahapan penting sebelum data digunakan untuk membangun model prediksi. Pada tahap *pre-processing* dilakukan proses pengecekan *missing value*, duplikasi data, pengecekan *outlier* dan penentuan atribut yang akan digunakan dalam pembuatan model prediksi. Berikut ini adalah proses yang dilakukan pada tahap *pre-processing*.

a. Pengecekan *missing value*

Pada proses pengecekan *missing value* pada *dataset* yang digunakan ditemukan beberapa variabel yang memiliki *missing value* seperti terlihat pada Tabel 2.

Tabel 2. Pengecekan *Missing Value*

Variabel	Jumlah <i>Missing Value</i>	Presentasi
closed_at	588	63.71%
Unnamed: 6	493	53.41%
age_last_milestone_year	152	16.47%
age_first_milestone_year	152	16.47%
state_code.1	1	0.11%

Berdasarkan Tabel 2 ditemukan 5 variabel yang memiliki *missing value* dengan *closed_at* dan *Unnamed:6* memiliki persentase *missing value* yang paling banyak yaitu di atas 50%. *Closed_at* adalah variabel yang berisi tanggal tutupnya sebuah *startup* dan *Unnamed:6* berisi gabungan beberapa variabel yaitu *state_code*, *city* dan *zip_code*. Kedua variabel tersebut tidak mendukung dalam pembuatan model prediksi sehingga dilakukan penghapusan variabel. Selain kedua variabel tersebut, variabel *state_code.1* juga dihapus meskipun persentase *missing value*-nya kecil karena memiliki nilai yang sama dengan variabel *state_code*.

Untuk mengatasi *missing value* pada variabel *age_last_milestone_year* dan *age_first_milestone_year* maka dilakukan analisis terhadap kedua variabel tersebut. Variabel *age_first_milestone_year* memiliki nilai rata-rata sebesar 3.055 dengan median 2.520 dan *age_last_milestone_year* memiliki nilai rata-rata sebesar 4.754 dengan median 4,477 yang menunjukkan bahwa data dari kedua variabel tersebut terdistribusi tidak normal. Sehingga *missing value* pada kedua variabel tidak dapat diisi dengan nilai rata-rata. Kemudian dilakukan analisis kedua variabel dengan variabel *milestones* sehingga diperoleh hasil bahwa *startup* yang belum memiliki *milestones* atau nilai *milestones* = 0 maka seluruh variabel *age_last_milestone_year* dan *age_first_milestone_year* memiliki nilai *null (NaN)* seperti terlihat pada Tabel 3.

Tabel 3. Nilai Variabel *age first milestone year*, *age last milestone year* dan *milestones*

No	<i>age first milestone year</i>	<i>age last milestone year</i>	<i>milestones</i>
1	4.6685	6.7041	3
2	7.0055	7.0055	1
3	1.4575	2.2055	2
4	6.0027	6.0027	1
5	0.0384	0.0384	1
...
95	NaN	NaN	0
96	3.2849	4.9315	3
97	0.0000	0.0000	1
98	2.4164	5.5151	2
99	NaN	NaN	NaN

Berdasarkan hasil analisis pada Tabel 3, untuk mengatasi seluruh *missing value* pada variabel *age first milestone year* dan *age last milestone year* dilakukan dengan teknik *imputation*. Salah satu cara untuk melakukan *imputation* adalah dengan memberikan nilai konstanta yang merupakan bagian dari kumpulan nilai yang mungkin dari variabel yang mengalami *missing value*. Pada penelitian ini nilai konstanta yang digunakan adalah angka 0 yang artinya *startup* tersebut belum memiliki *milestones* sama sekali selama beroperasi.

b. Pengecekan *duplicate value*

Setelah dilakukan proses penanganan *missing value* selanjutnya dilakukan pengecekan apakah ada data yang *duplicate*. Hasil pengecekan menunjukkan hanya ada 1 data yang *duplicate* seperti terlihat pada Tabel 4 dan pada tahap ini dilakukan proses penghapusan data *duplicate*.

Tabel 4 *Duplicate Value*

No	Unnamed:0	state_code	latitude	longitude	zip_code	id	...	status
832	505	CA	37.48151	- 121.945328	94538	C:28482	...	acquired

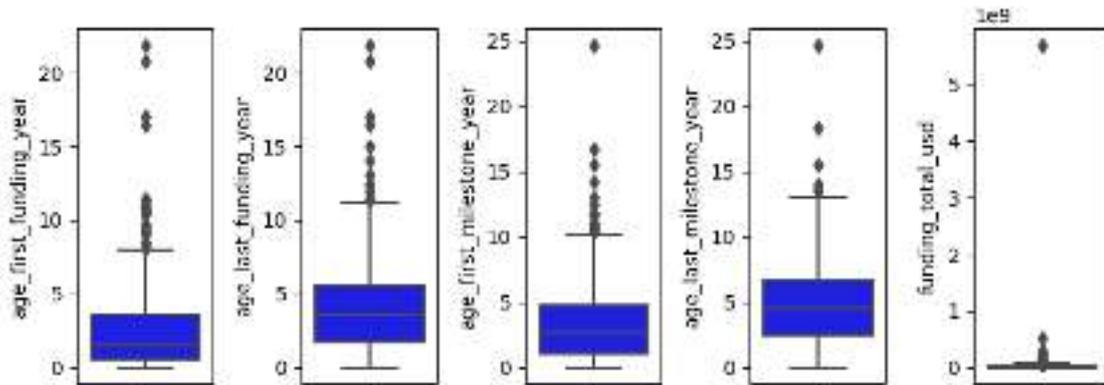
c. Pengecekan *outlier*

Sebelum melakukan pengecekan *outlier* pada data, maka dilakukan proses pengecekan nilai negatif pada beberapa *variabel* numerik dan ditemukan beberapa variabel memiliki nilai negatif seperti terlihat pada Tabel 5. *Negative value* pada data tersebut kemudian dihapus [5] sehingga tidak ada lagi data yang bernilai negatif.

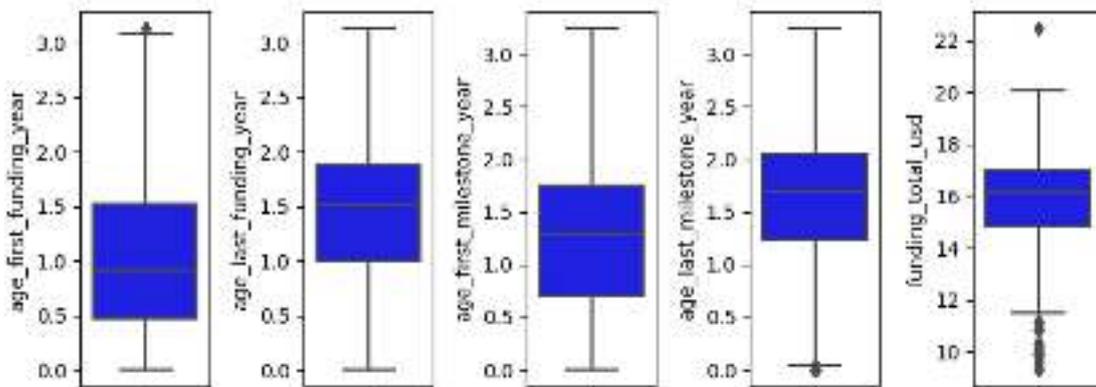
Tabel 5. *Negative Value*

No	Variabel	Negative Value
1	<i>age first milestone year</i>	46
2	<i>age last milestone year</i>	12
3	<i>age first funding year</i>	46
4	<i>age last funding year</i>	12

Selanjutnya dilakukan proses pengecekan *outlier* dan ditemukan beberapa variabel yang memiliki *outlier* seperti pada variabel *age first funding year*, *age last funding year*, *age first milestone year*, *age last milestone year*, *funding total usd*. Gambar 4 merupakan grafik *outlier* yang ditemukan pada data.

Gambar 4. *Outlier Pada Data*

Untuk mengatasi *outlier* pada data maka digunakan fungsi *log-transformation* untuk setiap variabel [5], sehingga data dapat terdistribusi normal seperti terlihat pada Gambar 5.

Gambar 5. Hasil *Log-Transformation Pada Outlier Data*

d. Penentuan atribut/variabel

Pada tahap ini dilakukan penentuan atribut/variabel yang akan digunakan untuk pembuatan model prediksi. Untuk menentukan variabel yang akan digunakan maka dilakukan proses hubungan atau korelasi antar variabel. Korelasi variabel direpresentasikan dengan nilai +1 sampai -1. +1 menunjukkan korelasi positif tertinggi antar variabel, 0 menunjukkan tidak ada korelasi antar variabel dan -1 menunjukkan korelasi negatif tertinggi antar variabel.

Berdasarkan hasil analisis terhadap korelasi 49 variabel terdapat 32 variabel yang dinilai memiliki pengaruh signifikan apakah sebuah *startup* akan sukses atau tidak. 32 variabel tersebut adalah *age_first_funding_year*, *age_last_funding_year*, *age_first_milestone_year*, *age_last_milestone_year*, *relationships*, *funding_rounds*, *funding_total_usd*, *milestones*, *is_CA*, *is_NY*, *is_MA*, *is_TX*, *is_otherstate*, *is_software*, *is_web*, *is_mobile*, *is_enterprise*, *is_advertising*, *is_gamesvideo*, *is_ecommerce*, *is_biotech*, *is_consulting*, *is_othercategory*, *has_VC*, *has_angel*, *has_roundA*, *has_roundB*, *has_roundC*, *has_roundD*, *avg_participants*, *is_top500*, *status*. Pada penelitian ini juga ditambahkan 4 variabel tambahan yaitu *has_RoundABCD*, *has_VC*, *has_Seed* dan *age_startup_year* sesuai dengan penelitian yang dilakukan oleh Goldenia *at al*. Dengan keempat variabel tersebut dapat mendukung model prediksi yang akan dibuat, sehingga total variabel yang digunakan untuk pembuatan model prediksi ada 36.

3.3 Pembagian Dataset

Pada tahap ini dilakukan proses pembagian *dataset* menjadi 80% *data training* dan 20% *data testing*. Proses pembagian *dataset* dilakukan secara *random* [5]. Hasil Pembagian *dataset* diperoleh

sebanyak 672 data sebagai *data training* dan 168 data sebagai *data testing* dengan rincian datanya dapat dilihat pada Tabel 6.

Tabel 6. Pembagian *Dataset*

Jenis Data	Status	
	<i>Acquired</i>	<i>Closed</i>
<i>Training</i>	433	239
<i>Testing</i>	119	49

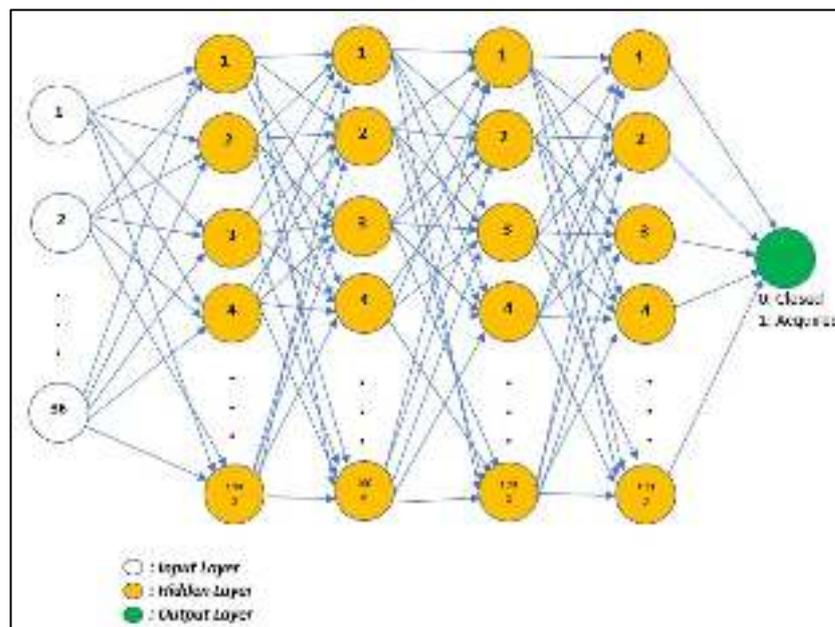
3.4 Pembuatan Model Prediksi

Pada tahap ini dilakukan pembuatan model prediksi menggunakan DNN. Tujuan pembuatan model prediksi menggunakan DNN adalah untuk menghasilkan akurasi yang lebih baik dan model tidak mengalami *underfitting* dan *overfitting*. Data yang digunakan untuk membangun model adalah *data training* sebanyak 672 data. Penentuan parameter ditentukan berdasarkan penelitian [7] dan melalui percobaan serta hasil analisis selama pembuatan model dilakukan.

Berikut parameter yang digunakan untuk membangun model menggunakan DNN.

- Input Layer* dengan jumlah variabel 36 sesuai dengan jumlah atribut yang digunakan sebagai *input*.
- 4 *Hidden layer* dengan jumlah 1000 *neuron*
- Output layer* dengan jumlah 1 *neuron*
- Dropout* = 0.03 dan *L2-regularization* = 0.01 untuk mengatasi terjadi *overfitting* pada model.
- Fungsi aktivasi yang digunakan adalah *Relu* dan *Sigmoid* karena model prediksi yang dibangun hanya untuk *binary classification*. *Relu* digunakan pada *hidden* dan *input layer* dan *Sigmoid* digunakan pada *output layer*.
- Fungsi optimasi yang digunakan adalah *Adam*. *Adam* menggabungkan sifat-sifat terbaik dari algoritma *AdaGrad* dan *RMSProp* untuk memberikan *optimization algorithm* yang dapat menangani *sparse gradients* pada *noisy problem*.
- Jumlah *epoch* adalah 300.
- Metrik pengukuran yang digunakan adalah *accuracy*, *precision*, dan *error*.
- Loss function* yang akan digunakan adalah *binary_crossentropy* karena data yang digunakan hanya memiliki *output* 0 dan 1 dan pada fungsi aktivasi *sigmoid* yang digunakan pada *output layer* juga mempunyai nilai 0 sampai 1.

Arsitektur model prediksi yang diusulkan dapat dilihat pada Gambar 6.

Gambar 6. Arsitektur Model Prediksi *Deep Neural Network*

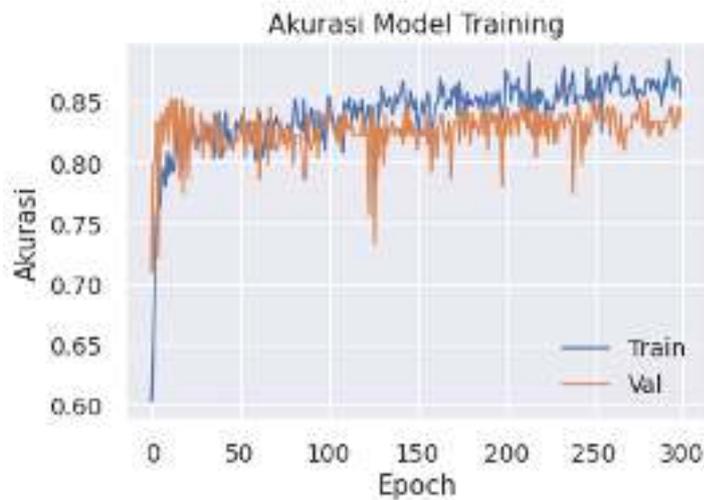
3.5 Evaluasi Model

Model prediksi yang sudah dibangun menggunakan data *training* kemudian dievaluasi menggunakan *confusion matrix* untuk menghasilkan nilai akurasi dan presisi berdasarkan 168 data *testing* yang digunakan untuk proses evaluasi model.

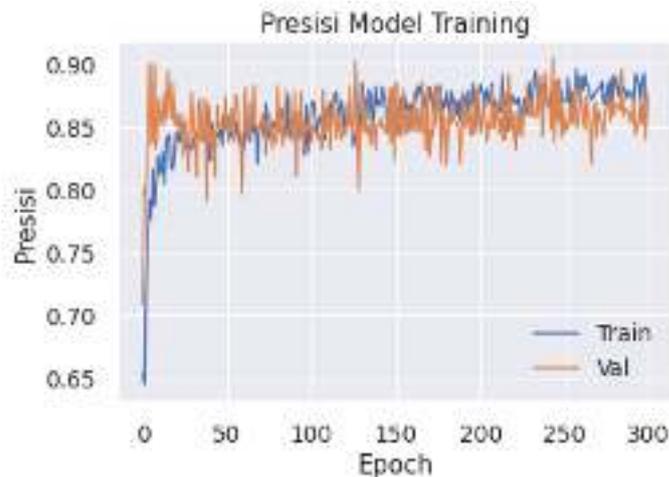
4. HASIL DAN PEMBAHASAN

4.1 Hasil Pembuatan Model Prediksi

Dataset yang digunakan memiliki dua kelas yaitu *Acquired* (Sukses) dan *Closed* (Gagal). Gambar 7 dan 8 merupakan grafik nilai akurasi dan presisi yang didapatkan pada proses pembuatan model prediksi.



Gambar 7. Akurasi Model Prediksi *Training*

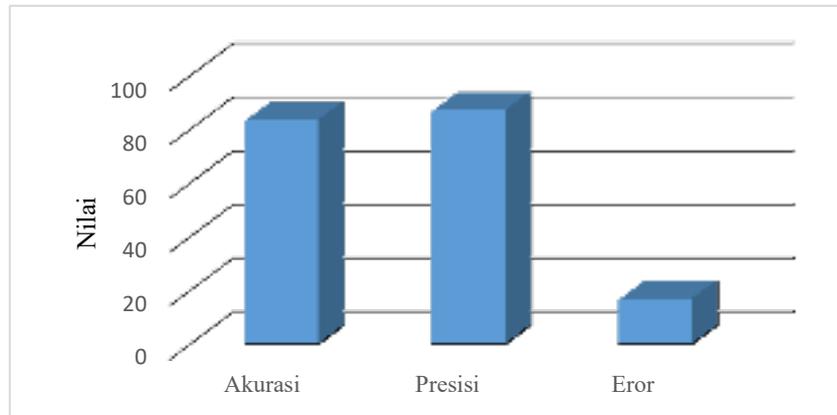


Gambar 8. Presisi Model Prediksi *Training*

Berdasarkan Gambar 7 dan Gambar 8 maka model prediksi yang dibangun tidak *underfitting* ataupun *overfitting* sehingga model dapat digunakan untuk proses prediksi kesuksesan sebuah *startup* dengan nilai akurasi sebesar 89,29%, presisi sebesar 90,20% dan eror 10,71%.

4.2 Evaluasi Model Prediksi

Model prediksi yang sudah dibangun menggunakan *data training* kemudian dievaluasi menggunakan *data testing* dan menghasilkan nilai akurasi sebesar 83,33%, nilai presisi sebesar 86,99% dan eror sebesar 16,07%. Gambar 9 merupakan hasil dari nilai akurasi, presisi dan eror dari model prediksi yang dibangun.



Gambar 9. Hasil Evaluasi Model Prediksi

Model prediksi yang sudah dibangun kemudian dapat digunakan untuk membantu para investor memprediksi apakah sebuah *startup* akan sukses atau tidak. Tabel 7 merupakan hasil penggunaan model prediksi dengan 10 data *testing* yang digunakan dalam prediksi kesuksesan *startup* ditemukan bahwa model dapat memprediksi 9 data *testing* dengan benar.

Tabel 7. Hasil Prediksi Kesuksesan Startup

Data Testing	Status Aktual	Hasil Prediksi Kesuksesan dalam %
1	<i>Acquired</i>	92.14
2	<i>Acquired</i>	89.23
3	<i>Closed</i>	9.52
4	<i>Closed</i>	25.50
5	<i>Closed</i>	48.0
6	<i>Closed</i>	93.96
7	<i>Acquired</i>	92.72
8	<i>Acquired</i>	94.86
9	<i>Acquired</i>	96.20
10	<i>Acquired</i>	94.78

5. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, disimpulkan bahwa model prediksi kesuksesan *startup* menggunakan *Deep Neural Network* yang diusulkan lebih baik dibandingkan dengan penelitian sebelumnya [5] yang menggunakan *Support Vector Machine* dengan nilai akurasi sebesar 89.29% dan presisi sebesar 89.85% untuk tahap *training* dan nilai akurasi sebesar 83.93 % dan presisi sebesar 86.51% untuk tahap *testing*. Berdasarkan hasil tersebut, model prediksi yang sudah dibangun kemudian dapat digunakan untuk membantu para investor memprediksi apakah sebuah *startup* akan sukses atau tidak.

6. SARAN

Untuk penelitian selanjutnya model ini dapat diterapkan pada *dataset startup* pada kategori tertentu saja seperti yang pada penelitian [15] untuk menguji apakah model prediksi yang dibangun dapat menghasilkan nilai akurasi yang tinggi atau tidak. Selain itu, model prediksi yang dibangun juga dapat diuji dengan menggunakan *dataset* yang lain seperti *dataset* dari *angelist.co* dan lain-lain untuk

memprediksi kinerja sebuah *startup*. Penelitian selanjutnya juga dapat dilakukan analisis terhadap faktor atau atribut apa saja yang paling mempengaruhi kesuksesan sebuah *startup*.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Mikroskil yang telah memberi dukungan finansial terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] Birkman, Yevgeniy, "Hello, Startup: A Programmer's Guide to Building Products, Technologies, and Teams", O'Reilly Media, Inc., 2015.
- [2] Li, Jinze, "Prediction of the Success of Startup Companies Based on Support Vector Machine and Random Forest", 2nd International Workshop on Artificial Intelligence and Education (WAIE 2020), Montreal, QC, Canada. ACM, New York, November 2020.
- [3] Ramalakshmi, Eliganti, and Kamidi, Sindhuja Reddy, "Predictions for Startups", International Journal of Engineering & Technology, Vol 7, No 3.12, 2018.
- [4] Permana, et al., "Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up", JISKA, Vol. 6, No. 3, Pp. 178 – 188, September 2021.
- [5] Goldenia, et al., "Implementasi Algoritma Support Vector Machine dalam Memprediksi Keberhasilan Suatu Startup Berdasarkan Status Akuisisi", Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA), September 2021.
- [6] Yu, Pengfei and Yan, Xuesong, "Stock Price Prediction based on Deep Neural Networks", Neural Computing and Applications 32, pp. 1609–1628, 2020.
- [7] Faisal, Anas and Subekti, Agus, "Deep Neural Network untuk Prediksi Stroke", Jurnal Edukasi dan Penelitian Informatika, Vol 7, No 3, Desember 2021.
- [8] Feng et al., "Using Deep Neural Network with Small Dataset to Predict Material Defects", Materials and Design Elsevier, Vol 162, pp. 300-310, Januari 2019.
- [9] Rahardjo, Budi, "Starting Up", PT Insan Indonesia, 2020.
- [10] Hardiansyah, Rudi, and Tricahyono, Dodie, "Identifikasi Faktor-Faktor Kesuksesan Start Up Digital di Kota Bandung", Jurnal Ekonomi, Vol. 27, No. 2, Pp. 134-145, 2019
- [11] Jayen Suwarno, Dodi, and Silvianita, Anita, "Knowledge Sharing Dan Inovasi Pada Industri Startup", Jurnal Ecodemica, Vol. 1, No. 1, April 2017.
- [12] Seasia.co. "As internet economy grows in Southeast Asia, more unicorns to emerge" Seasia, 1 Nov. 2021, [Online] Tersedia: <https://seasia.co/2021/11/01/as-internet-economy-grows-in-southeast-asia-more-unicorns-to-emerge> [Diakses: 19 September 2022]
- [13] Yuvaraj, et. al., "Analysis on the prediction of central line-associated bloodstream infections (CLABSI) using deep neural network classification", Computational Intelligence and Its Applications in Healthcare, ScienceDirect, Academic Press, pp. 229 - 244, 2020.
- [14] KC, Manish, "Startup Success Prediction", kaggle, September 2020, [Online], Tersedia: <https://www.kaggle.com/datasets/manishkc06/startup-success-prediction> [Diakses: 1 Juli 2022]
- [15] K T, Salsabila Amnes and Herianto, "Perancangan Model Prediksi Kesuksesan Startup 3D printing Menggunakan Logistic Regression", TALENTA Conference Series: Energy and Engineering (EE), Vol. 4 Issue 1, 2021.