

ANALISA CARA KERJA MICROSERVICE BERBASIS PHYTON UNTUK PERANCANGAN CREDIT SCORE PADA DI FINTECH

Andre Pratama Adiwijaya¹, Sutresna Wati²

¹ Fakultas Ekonomi, Andre64@staff.gunadarma.ac.id, Universitas Gunadarma

² Direktorat DIII Teknologi Informasi, tresna@staff.gunadarma.ac.id, Universitas Gunadarma

ABSTRACT

This research is able to determine the feasibility of providing loans online without having to conduct an offline feasibility survey. The modeling for the feasibility of giving uses the Predictive Analytics model scheme where this service/application can stand alone or join the entire application of ERP (Enterprise Resource Planning). The method used in designing services/applications based on agile. This system is based on Python and also a database using MYSQL. This application makes it easy for financial companies from multi-finance to banks to choose prospective borrowers with low default rates and make the TKB value of the company always maintained. And can minimize borrowers default.

Keywords: Fintech, Online Loan Assesment, Phyton, Predictive Analytics.

ABSTRAK

Penelitian ini mampu membuat penentuan kelayakan pemberian pinjaman secara online tanpa harus melakukan survei kelayakan secara offline. Permodelan untuk kelayakan pemberian menggunakan skema model Predictive Analytics yang dimana service/aplikasi ini bisa berdiri sendiri ataupun bergabung dengan keseluruhan aplikasi dari ERP (Enterprise Resource Planning). Metode yang digunakan dalam perancangan pembuatan service/aplikasi berisikan agile. Sistem ini berbasiskan Phyton dan juga database menggunakan MYSQL. Aplikasi ini memudahkan perusahaan finansial dari multi finance hingga bank dapat memilih calon peminjam dengan tingkat gagal bayar rendah serta membuat nilai TKB dari perusahaan tersebut akan selalu terjaga. Dan dapat meminimalisir para peminjam gagal bayar.

Kata Kunci: Predictive Analytics, Fintech, Online Loan Assesment , Phyton.

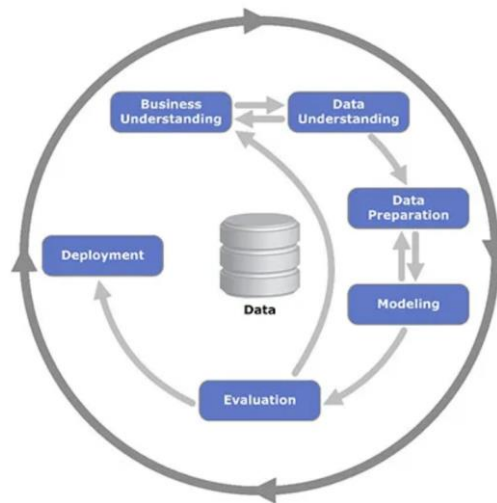
1. PENDAHULUAN

Pada saat ini zaman mengalami perkembangan teknologi yang sangat pesat. Setiap harinya teknologi semakin berkembang. Saat ini penilaian terhadap kredibilitas dari peminjam secara individual dan korporasi sangat manual dan secara offline. Perancangan sistem loan assesment ini berbasiskan web, para perusahaan finansial dapat memberikan penelitian terhadap calon nasabah bisa dilakukan secara online dengan parameter yang disupport dari berbagai sumber yang akan dijadikan satu model oleh system yang menggunakan metode Predictive Analytic. Dimana hasil permodelan tersebut akan menghasilkan nilai A sampai E tergantung dari calon nasabah yang akan dinilai. Aplikasi menggunakan Visual Studi Code dengan tiga buah Bahasa pemrograman yaitu Phyton. Metode Predictive Analytic merupakan salah satu metode untuk melakukan prediksi dari suatu rumusan dari berbagai sumber yang di jadikan suatu score kelayakan pemberian pinjaman. Supaya tepat sasaran dan juga tidak memberatkan saat melakukan penagihan yang membuat nilai dari TKB atau tingkat keberhasilan penyaluran pinjaman.

2. TINJAUAN PUSTAKA

2.1. *Predictive Analytic*

Predictive analytics merupakan cabang dari *advanced analytics* yang pada proses kerjanya menggunakan prediksi guna mengetahui kejadian di masa yang akan datang. *Predictive analytics* juga menggunakan banyak teknik mulai dari data mining, statistik, *machine learning* untuk menganalisis data saat ini dan membuat prediksi untuk kondisi di masa mendatang. Dalam dunia bisnis, hal ini dapat membantu menemukan pola transaksi terdahulu untuk mengidentifikasi risiko dan peluang. Sifat dari *predictive analytic* juga dapat mengkategorikan dari beberapa transaksi ataupun data terkait keuangan sebelumnya yang membuat proses penilaian calon nasabah bisa dilakukan secara proporsional. Adapun tahapan pembuatan untuk melakukan *modelling predictive analytic* dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Pembuatan Modelling menggunakan *Predictive Analytic*

2.2. Teknik Pengumpulan Data

Teknik Pengumpulan Data yang dilakukan diantaranya :

1. Studi pustaka, membaca dan mencari artikel serta jurnal terkait *credit assement*, *machine learning*, dan *phyton*;
2. Studi lapangan. Teknik ini digunakan untuk melakukan pembuatan permodelan dasar untuk menghitung penilain kelayakan kredit terhadap nasabah.

2.3. Phyton

Phyton merupakan bahasa pemrograman yang berbasis OOP yang mampu mengimplementasikan arsitektur MVC (model view controller). MVC merupakan sebuah pendekatan perangkat lunak yang memisahkan aplikasi logika dari presentasi. MVC memisahkan aplikasi berdasarkan komponen-komponen aplikasi, seperti : manipulasi data, controller, dan user interface.

- Model, Model mewakili struktur data. Biasanya model berisi fungsi-fungsi yang membantu seseorang dalam pengelolaan basis data seperti memasukkan data ke basis data, pembaruan datadan lain-lain.
- View, View adalah bagian yang mengatur tampilan ke pengguna. Bisa dikatakan berupa halamanweb.

2.4 Dataset

Dataset digunakan dalam pembuatan aplikasi ini digunakan sebagai sumber data yang dijadikan acuan ataupun sumber perbandingan data dari nasabah yang mengajukan pinjaman ke institusi keuangan seperti *fintech*, *multifinance* dan bank. Adapun tipe file yang digunakan di dalam database ini adalah csv/xls yang akan dibuatkan sebagai parameter untuk penilaian terhadap nasabah.

3. METODELOGI PENELITIAN

Aplikasi *Loan Assement* adalah aplikasi berbasis website yang dapat diakses dari manapun dan kapanpun. Proses pembuatan data modeling pada Aplikasi Loan Assesment dapat digunakan oleh aplikasi lain dengan cara memanfaatkan API (*Application Programming Interface*) dengan protokol komunikasi menggunakan *JWT 2* sebagai proses keamanan pertukaran data. Aplikasi Loan Assesment menggunakan beberapa data set agar hasil penilaian mendapatkan calon nasabah dengan kredibilitas yang baik terkait pembayaran hutang, mengakitbatkan nilai dari TKB suatu perusahaan finansial bisa terjaga dan uang dari investor dapat dikembalikan tanpa harus melakukan *claim* asuransi. Aplikasi Loan Assesment menggunakan service *credit scoring* untuk membantu memvalidasi data pendapatan atau kemampuan bayar dari seorang calon nasabah. Data yang dikumpulkan dari *credit scoring* lalu dibuat beberapa model calon nasabah dengan menggunakan metode *predictive analytic*. *Predictive analytic* juga dapat langsung menghasilkan beberapa model calon nasabah dengan menggunakan metode regresi.

Di dalam aplikasi ini terdapat 32.581 peminjam dengan 11 variable yang terkait dengan peminjam berikut variabel yang digunakan dalam aplikasi ini :

- Age — numerical variable; age in years
- Income — numerical variable; annual income in dollars
- Home status — categorical variable; “rent”, “mortgage” or “own”
- Employment length — numerical variable; employment length in years
- Loan intent — categorical variable; “education”, “medical”, “venture”, “home improvement”, “personal” or “debt consolidation”
- Loan amount — numerical variable; loan amount in dollars
- Loan grade — categorical variable; “A”, “B”, “C”, “D”, “E”, “F” or “G”
- Interest rate — numerical variable; interest rate in percentage
- Loan to income ratio — numerical variable; between 0 and 1
- Historical default — binary, categorical variable; “Y” or “N”
- Loan status — binary, numerical variable; 0 (no default) or 1 (default) → this is going to be our target variable

3.1 Implementasi

Setelah mendapatkan dataset sample berupa 11 variable dari 32.581 peminjam maka hal yang pertama akan dilakukan pengecekan data seperti gambar 3.1.

```
#Checking for missing values
data.isnull().sum()

Age                0
Income             0
Home_Status        0
Employment_Length  895
Loan_Intent        0
loan_Grade         0
Loan_Amount        0
Interest_Rate      3116
Loan_Status        0
loan_percent_income 0
Historical_Default 0
dtype: int64
```

Gambar 2. Cheking Value

Selanjutnya dalam aplikasi ini dilihat lama bekerja dengan tingkat bunga pinjaman, tetapi untuk mendapatkan hasil yang lebih proposional maka sebagian kecil dari kumpulan data, akan menghapus baris nilai yang hilang. Untuk *syntax* yang digunakan adalah `#Dropping missing values data = data.dropna(axis=0)`, kemudian cari *outlier* di dataset sehingga dapat diperbaiki dengan tepat.

Tahapan selanjutnya menggunakan metode `explain()` yang digunakan untuk menghitung statistik deskriptif. Tidak hanya akan membantu mengidentifikasi *outlier*, tetapi juga akan memberikan pemahaman yang lebih baik mengenai pendistribusian data. Untuk membangun aplikasi ini digunakan juga *matriks scatterplot*, kisi-kisi *scatterplot* yang digunakan untuk memvisualisasikan hubungan *bivariat* antara kombinasi variabel, untuk mendeteksi *outlier* secara visual. Untuk menampilkan data tersebut menggunakan *syntax* `data.describe()`. Data akan muncul seperti pada gambar 3.2 yang sudah diolah sebelumnya.

	Age	Income	Employment_Length	Loan_Amount	Interest_Rate	Loan_Status	loan_percent_income
count	28638.000000	2.863800e+04	28638.000000	28638.000000	28638.000000	28638.000000	28638.000000
mean	27.727216	6.664937e+04	4.784482	9656.493121	11.039867	0.216600	0.169488
std	6.310441	6.235645e+04	4.095491	6329.683361	3.229372	0.411935	0.106393
min	20.000000	4.000000e+03	0.000000	500.000000	5.420000	0.000000	0.000000
25%	23.000000	3.948000e+04	2.000000	5000.000000	7.900000	0.000000	0.090000
50%	26.000000	5.595600e+04	4.000000	8000.000000	10.990000	0.000000	0.150000
75%	30.000000	8.000000e+04	7.000000	12500.000000	13.480000	0.000000	0.230000
max	144.000000	6.000000e+06	123.000000	35000.000000	23.220000	1.000000	0.830000

Gambar 3. Dataset yang sudah diolah

Terdapat beberapa kondisi anomali yang muncul seperti banyak orang yang telah hidup sampai umur usia 144 atau telah bekerja selama 123 tahun. Selain itu, *outlier* dari variabel usia dan lama kerja dapat berdampak negatif pada model, oleh karena itu, harus dihilangkan. Selanjutnya, perlu mencari lebih banyak *outlier* menggunakan *matriks scatterplot*. Berikut *syntax* untuk pembuatan *matrix scatterplot* untuk membuat normalisasi umur sampai bunga terdapat pada gambar 3.3

```
#Scatterplot matrix
fig = px.scatter_matrix(data, dimensions=
["Age","Income","Employment_Length","Loan_Amount","Interest_Rate"],
labels={col:col.replace('_', ' ') for col in data.columns},
height=900, color="Loan_Status",
color_continuous_scale=px.colors.diverging.Tealrose)
fig.show()
```

Gambar 4. *Syntax matrix scatterplot*



Gambar 5. Bagan *scatterplot*

Melalui pembuatan *matrix scatterplot* akan didapatkan *modeling* data dari dataset yang sebelumnya yang berbebetuk bagan *scatterplot*. Yang mana nilai dari bagan sudah bagus dan dapat terlihat jelas titik – titik dari dari dataset dan dapat dilihat pada Gambar 3.4.

3.2 Pembersihan Outlier

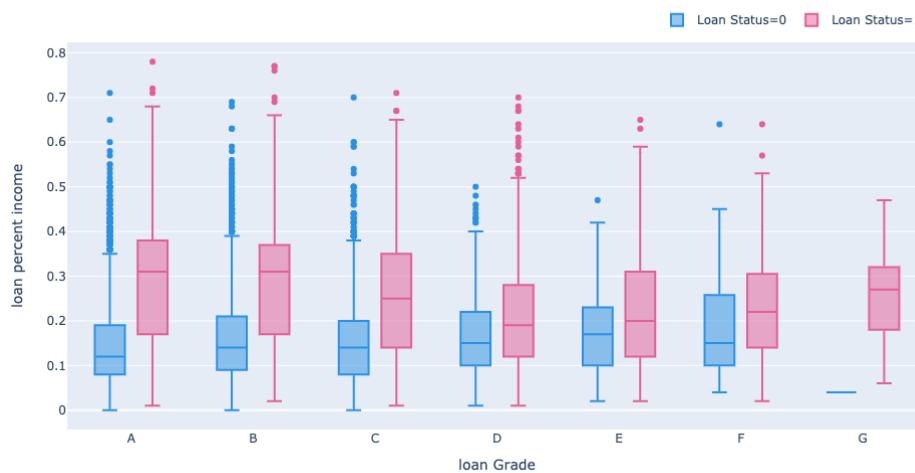
Pada dataset yang sudah ditampilkan pada gambar 3.4 menggunakan bagan *scatterplot*, dapat ditarik kesimpulan sementara *modeling* data berdasarkan umur, *income*, bunga dan jumlah pinjaman. Akan tetapi masih ada *Outlier* yang masih akan dirapihkan kembali yaitu:

- `data = data[data["Age"]<=100]`
- `data = data[data["Employment_Length"]<=100]`
- `data = data[data["Income"]<= 4000000]`

Dari kumpulan dataset yang digunakan dalam aplikasi ini ada beberapa data yang tidak seimbang mengakibatkan nilai dari beberapa data tidak cukup bagus dari segi klasifikasi. Dalam aplikasi terdapat kasus terhadap dataset, banyak yang melebihi dari standar atau default maka dari itu untuk meminimalisir hal tersebut, diperlukan mengeluarkan *Outlier* data menggunakan *syntax*:

- #Percentage of non-default cases
- data_0 = data[data.Loan_Status == 0].
- Loan_Status.count()/data.Loan_Status.count() data_0

Untuk mengkonfirmasi bahwa 78,4% dari kumpulan data set berisi kasus non-default. Dataset akan dieksplorasi lebih lanjut bagaimana status pinjaman terkait dengan variabel lain dalam kumpulan dataset. Untuk menampilkan data yang sudah dieksplorasi di dalam aplikasi, digunakan pemodelan visualisasi berbentuk *square plot*, agar dapat terlihat jelas bawa nasabah yang tidak gagal bayar memiliki resiko rasio pinjaman terhadap pendapatan yang lebih rendah berarti nilai di semua peringkat pinjaman yang tidak mengejutkan. Dalam aplikasi ini juga dapat melihat bahwa tidak ada peminjam dengan peringkat G yang tidak membayar kembali pinjaman mereka. Dengan menggunakan diagram kategori paralel, dapat dipahami bagaimana variabel yang tidak berbeda dalam dataset terkait satu sama lain dan dapat memetakan hubungan variabel berdasarkan status pinjaman dapat dilihat pada gambar 3.5.

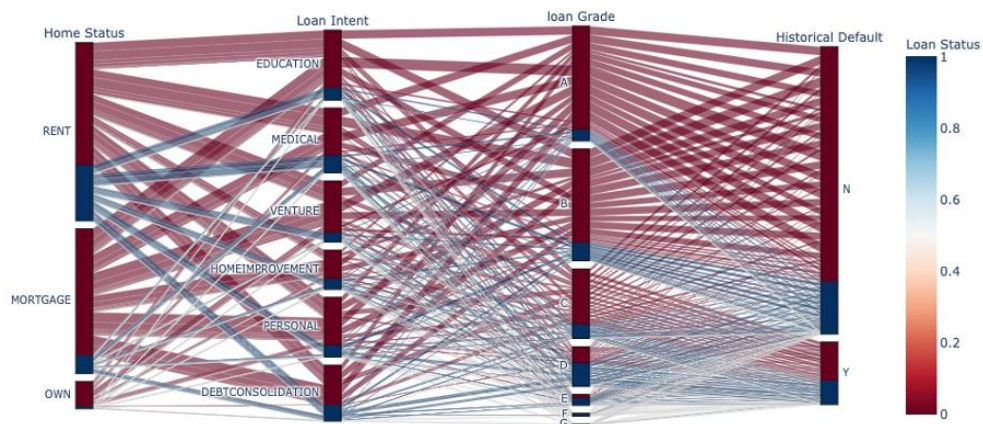


Gambar 6. *Square Plot Diagram*

Dataset yang sudah dilakukan eksplorasi dengan menggunakan *square plot*, dilakukan pengembangan dari dataset menggunakan diagram kategori yang bertujuan memisahkan nilai pinjaman menjadi beberapa *grade*/tingkatan seperti pada gambar 3.6 yang memiliki beberapa tingkatan nilai peminjam. Terdapat beberapa hal yang harus diperhatikan seperti di bawah ini:

- Dataset terdiri dari peminjam yang belum pernah gagal membayar pinjaman sebelumnya.
- Nilai pinjaman "A" dan "B" adalah nilai yang paling umum sementara "F" dan "G" adalah yang paling tidak umum.
- Penyewa rumah lebih sering gagal membayar pinjaman mereka daripada mereka yang memiliki hipotek, sedangkan pemilik rumah paling sedikit gagal.
- Peminjam mengambil pinjaman untuk perbaikan rumah paling sedikit dan untuk pendidikan paling banyak. Juga, default lebih umum untuk pinjaman yang diambil untuk menutupi biaya pengobatan dan konsolidasi utang.

Sebelum masuk ke pelatihan model, perlu memastikan bahwa semua variabel adalah numerik mengingat beberapa model yang akan digunakan tidak dapat beroperasi pada data label. Cukup melakukan ini dengan menggunakan metode *one-hot encoding*.



Gambar 7. Diagram Kategori

3.3 Pelatihan Modeling dan Evaluasi

Pada bagian ini, akan melatih dan menguji 3 model, yaitu KNN, regresi logistik dan XGBoost. Di dalam aplikasi ini juga akan mengevaluasi kinerja mereka dalam memprediksi default pinjaman dan probabilitasnya. Pertama, yang akan membangun model dan melihat beberapa metrik evaluasi untuk menilai kemampuan model dalam memprediksi label kelas, yaitu *default* atau tanpa *default*.

Di dalam aplikasi ini telah mengidentifikasi sebelumnya bahwa berurusan dengan *dataset* yang tidak seimbang sehingga perlu memastikan bahwa menggunakan metrik evaluasi yang sesuai untuk kasus ini.. Untuk mengilustrasikan mengapa hal ini terjadi, akurasi menghitung rasio total nilai prediksi yang sebenarnya terhadap jumlah total sampel input, yang berarti bahwa model akan mendapatkan akurasi yang cukup tinggi dengan memprediksi kelas mayoritas tetapi akan gagal menangkap kelas minoritas, default, yang bukan *buena*. Inilah sebabnya mengapa metrik evaluasi yang akan fokuskan untuk menilai kinerja klasifikasi model kami adalah Precision, Recall, dan skor F1.

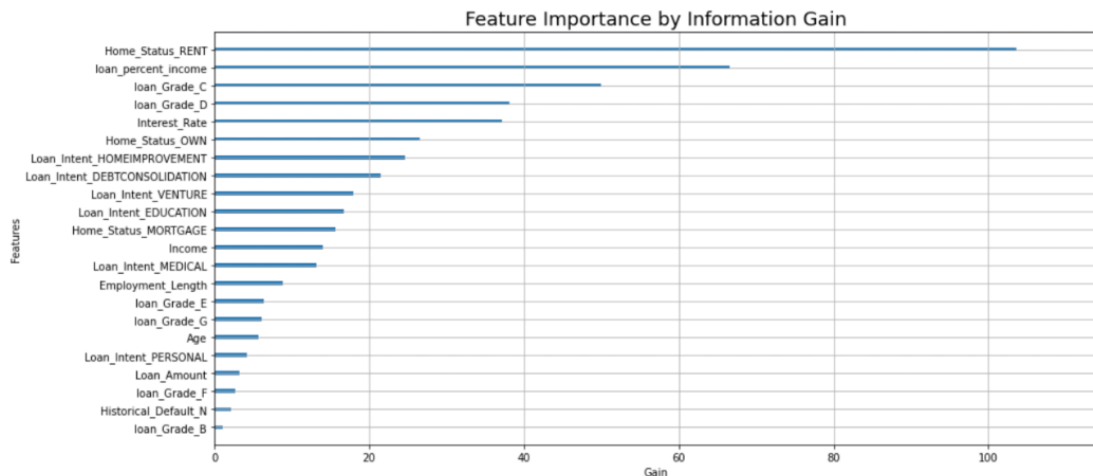
Pertama, Precision memberi rasio positif sejati dengan total positif yang diprediksi oleh pengklasifikasi di mana positif menunjukkan kasus default dalam konteks. Mengingat bahwa mereka adalah kelas minoritas dalam kumpulan data, dapat melihat bahwa hasil permodelan melakukan pekerjaan yang baik dalam memprediksi *instance minor* tersebut dengan benar. Selain itu, Recall, alias tingkat positif sejati, memberi jumlah positif sejati dibagi dengan jumlah total elemen yang sebenarnya termasuk dalam kelas positif. Dalam kasus ini, Recall adalah metrik yang lebih penting dibandingkan dengan Precision mengingat bahwa lebih peduli dengan negatif palsu (model kami memprediksi bahwa seseorang tidak akan default tetapi mereka melakukannya) dari pada positif palsu (model kami memprediksi bahwa seseorang akan default tapi mereka tidak). Terakhir, F1 Score memberikan skor tunggal untuk mengukur Precision dan Recall. Sekarang dapat diketahui apa yang harus dicari, dapat dengan jelas melihat bahwa XGboost melakukan yang terbaik di semua 3 metrik. Meskipun mencetak lebih baik pada Precision dibandingkan dengan Recall, ia masih memiliki skor F1 yang cukup bagus yaitu 0,81.

Sekarang akan melihat ROC yang merupakan kurva probabilitas dengan *False Positive Rate* (FPR) pada sumbu x dan *True Positive Rate* (TPR, recall) pada sumbu y. Model terbaik harus memaksimalkan TPR ke 1 dan meminimalkan FPR ke 0. Dengan ini, dapat membandingkan pengklasifikasi menggunakan area di

bawah kurva kurva ROC, AUC, di mana semakin tinggi nilainya, semakin baik model dalam memprediksi 0s sebagai 0s dan 1s sebagai 1s.

4. KESIMPULAN DAN SARAN

Hasil dari penulisan ini dapat terlihat fitur yang paling berpengaruh dalam memprediksi variabel target aplikasi *loan assement*. Untuk aplkasi ini akan menggunakan fitur penting dengan perolehan informasi yang mengukur kontribusi setiap fitur untuk setiap tree di *XGBoost*. Dapat terlihat pada gambar 4.1 bahwa sebagai status rumah, rasio pinjaman terhadap pendapatan dan pinjaman kelas C adalah 3 fitur terpenting untuk memprediksi default pinjaman dan probabilitasnya.



Gambar 8. Diagram Batang hasil pengolahan dataset

Singkatnya, aplikasi ini telah menganalisis dan memproses data set sebelumnya, melatih dan mengevaluasi 3 model, namely KNN, regresi logistik, dan *XGBoost*, untuk kemampuannya memprediksi default pinjaman dan probabilitasnya.

Menggunakan *Precision*, *Recall*, *F1* dan *ROCAUC* untuk mengevaluasi kinerja model dalam memprediksi label kelas. Didalam aplikasi ini juga menggunakan metrik ini secara khusus dan membuang Akurasi karena kami berurusan dengan kumpulan data yang tidak seimbang. Juga menggunakan *Reliability Plot* dan *Brier Score* untuk menilai kalibrasi model. Setelah mengidentifikasi bahwa *XGBoost* berkinerja terbaik pada semua metrik, menyelidiki fitur mana yang paling penting untuk prediksi dari aplikasi menggunakan kepentingan fitur berdasarkan perolehan informasi. Dengan ini, Aplikasi dapat mengumpulkan demonstrasi tentang bagaimana pembelajaran mesin dapat diterapkan ke dunia penilaian risiko kredit.

DAFTAR PUSTAKA

Thomas W. Miller, MODELING TECHNIQUES IN PREDICTIVE ANALYTICS: BUSINESS PROBLEMS.
 Ciby Joseph, ADVANCED CREDIT RISK ANALYSIS AND MANAGEMENT
 Jason LZP. "A Short Introduction To Decision Trees". Di akses dari <https://medium.com/gitconnected/a-short-introduction-to-decision-trees-9481c36d2c52> tanggal 20 Juni 2022
 Shoumik Goswami. "BUILDING A VISUAL MODEL MONITORING SYSTEM FOR CREDIT RISK MODELS ". Di akses dari <https://towardsdatascience.com/building-a-pipeline-monitoring-system-for-credit-risk-models-db55f9a49f7e> tanggal 21 Juni 2022
 I Gusti Ngurah Narindra Mandalaa, Catharina Badra Nawangpalupia, Fransiscus, Rian Praktiktoa. "ASSESSING CREDIT RISK: AN APPLICATION OF DATA MINING IN A RURAL BANK ". Di akses dari <https://reader.elsevier.com/reader/sd/pii/S2212567112003553?token=9461F2576DD22>

67D

[2C50C91C073787334EA6E94F4080945360EC549B55FAB65093D6ACBB22DFA1F8056CB9E470397819&originRegion=eu-west-1&originCreation=20220803074104](#)

tanggal 24 Juni 2022