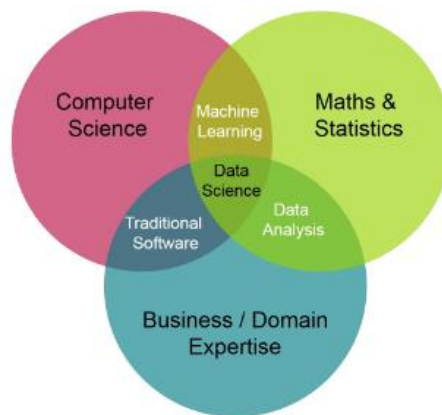


PERAN *DATA SCIENCE* DALAM *CREDIT SCORING* DI INDUSTRI PERBANKAN

PENDAHULUAN

Data Science adalah suatu ilmu yang mengombinasikan beberapa disiplin ilmu, seperti matematika, statistika, ilmu komputer, *artificial intelligence* (AI), dan analisa data, untuk mengolah *big data* menjadi suatu *insight* yang berguna. *Data Science* dapat membantu dalam memproses data yang meliputi pengumpulan data, manipulasi data, hingga analisis data dengan memodelkan kumpulan data untuk menghasilkan informasi berupa *insight* yang berguna dan bisa dijadikan pedoman dalam pengambilan keputusan di masa depan. *Data Science* mengolah *big data* yang mencakup data terstruktur maupun tidak terstruktur dan tidak hanya data numerik saja, tetapi juga data berupa suara, gambar, teks, dan sebagainya [1].



Gambar 1 Disiplin Ilmu dalam *Data Science*

Data Science sebenarnya sudah dimanfaatkan banyak perusahaan untuk membantu mengelola data yang mereka miliki. Mereka menyadari bahwa dari kumpulan data yang diolah dengan tepat akan menghasilkan informasi yang berguna bagi kemajuan perusahaan. *Data Science* mulai umum terdengar pada tahun 2018. Sejak saat itu, *Data Science* menjadi populer dan mulai banyak diterapkan di berbagai bidang karena kelebihan dan manfaat yang ditawarkannya. Misalnya dalam industri perbankan, *Data Science* digunakan untuk mendeteksi penipuan secara real-time. Dengan *Data Science*, perusahaan perbankan dapat membuat sebuah model yang mampu mengklasifikasikan mana transaksi yang diprediksi sebagai penipuan dan mana yang bukan. Selain itu, *Data Science* juga digunakan untuk membuat segmentasi nasabah agar perusahaan dapat menghemat biaya pemasaran dengan menyusun strategi pemasaran yang tepat sasaran dan sesuai dengan karakteristik nasabah. *Data Science* juga digunakan dalam industri perbankan untuk manajemen risiko [2]. Adapun alat yang digunakan dalam *Data Science* yaitu *machine learning*, *data mining*, *deep learning*, dan *artificial intelligence*.

Credit scoring merupakan salah satu *use case machine learning* yang biasanya digunakan oleh perusahaan *finance*, terutama bank, untuk menentukan apakah seorang nasabah layak diberikan pinjaman atau tidak. *Credit scoring* menghasilkan nilai dari 300-850 dimana semakin tinggi nilainya maka semakin baik nasabah mencari pemberi pinjaman yang potensial. Perusahaan pemberi pinjaman sering kali meninjau *credit scoring*, terutama saat akan memutuskan untuk mengubah suku bunga atau batas kredit pada kartu kredit nasabah [3]. Perhitungan *credit scoring*

dipengaruhi oleh *credit history* (histori *repayment* nasabah dan panjang histori kredit), *credit enquiries* (jumlah permintaan kredit yang dilakukan sebelumnya), dan *trade lines* (jumlah kartu kredit dan pinjaman yang aktif). Analisa *Credit scoring* memiliki tantangan tersendiri, yaitu tidak semua peminjam memiliki histori kredit yang signifikan untuk mendapat *credit score* yang baik. Selain itu, tantangan yang harus dihadapi perusahaan *finance* dalam *credit scoring* ini adalah tingginya persaingan di industri ini menjadikan besar kecilnya perusahaan tersebut berpengaruh terhadap keputusan calon peminjam dalam menentukan '*provider*'nya. Seorang peminjam dengan *credit scoring* yang bagus besar kemungkinannya tidak akan memilih perusahaan kecil untuk melakukan pinjaman [4].

Setiap perusahaan perbankan pasti memiliki tujuan untuk memaksimalkan profit yang didapatkan dari pinjaman seorang nasabah, begitu pula dengan bank "ABC". Selama menjalankan bisnisnya, bank "ABC" pernah mendapatkan profit dari pinjaman nasabah hingga sebesar \$100, tetapi bank juga pernah mendapatkan kerugian hingga mencapai \$500. Agar dapat memaksimalkan profit dan meminimalkan kerugian, bank "ABC" membangun suatu *in-house risk model* untuk membuat keputusan pinjaman untuk *subprime mortgages* (tipe pinjaman yang diberikan kepada individu dengan *credit score* rendah). Bank ini memiliki data historis nasabah terkait catatan biro kredit dan hasil pinjaman dengan kategori '*good loan*' dan '*bad loan*'. Karena hasil pinjaman bertipe kategorik biner maka metode *machine learning* yang tepat untuk membuat sebuah model prediksi dan menyelesaikan permasalahan di bank "ABC" ini adalah regresi logistik.

LANDASAN TEORI

1. *Machine Learning*

Machine learning merupakan salah satu cabang dari *artificial intelligence* berupa mesin yang dikembangkan agar mampu mempelajari data dan melakukan tugas-tugas tertentu tanpa arahan dari penggunanya. *Machine learning* dikembangkan berdasarkan disiplin ilmu seperti statistika, matematika, dan *data mining*. Istilah *machine learning* pertama kali dikemukakan oleh beberapa ilmuwan matematika seperti Adrien Marie Legendre, Thomas Bayes dan Andrey Markov pada tahun 1920-an dengan mengemukakan dasar-dasar *machine learning* dan konsepnya. Sejak saat itu ML banyak yang mengembangkan [5]. *Machine learning* memiliki banyak kegunaan, salah satunya di bidang bisnis, yaitu membantu menyelesaikan berbagai permasalahan bisnis yang kompleks, mewujudkan automasi bisnis, dan mempermudah proses intelegensi bisnis [6].

Machine learning memiliki beberapa teknik belajar, tetapi secara umum *machine learning* terbagi menjadi *supervised learning* dan *unsupervised learning*. *Supervised learning* merupakan teknik yang digunakan pada data yang memiliki informasi tertentu dan bisa diterapkan secara langsung dengan memberikan label tertentu. Teknik ini diharapkan mampu memberikan target terhadap *output* yang dilakukan dengan membandingkan pengalaman belajar di masa lalu [5]. Salah satu algoritma yang termasuk dalam *supervised learning* adalah klasifikasi, regresi, *random forest*, *time series forecasting*, dan *artificial neural network*. Sementara itu, *unsupervised learning* merupakan teknik yang digunakan pada data yang tidak memiliki informasi yang bisa diterapkan secara langsung. Teknik ini diharapkan mampu membantu menemukan struktur atau pola tersembunyi pada data yang tidak memiliki label [5]. Salah satu algoritma yang termasuk dalam *supervised learning* adalah *clustering*, deteksi anomali, *association rules*, dan *deep belief network*.

2. Regresi Logistik

Regresi logistik merupakan salah satu jenis *supervised learning* yang biasa digunakan untuk membuat sebuah model prediksi faktor-faktor yang berpengaruh terhadap variabel tertentu, sama halnya dengan regresi linier. Perbedaan regresi logistik dan regresi linier dapat dilihat dari variabel respon nya, dimana variabel respon pada regresi logistik berjenis kategorik yang nilainya adalah ya/tidak, benar/salah, sukses/gagal, ataupun dalam bentuk bilangan biner 0/1. Persamaan regresi logistik dituliskan sebagai berikut.

$$\pi(x) = \frac{\exp(g(x))}{1 + \exp(g(x))}$$

Dimana $\pi(x)$ adalah proporsi terjadinya suatu kejadian dan $g(x) = \beta_0 + \beta_1x_1 + \dots + \beta_px_p$ [7].

METODOLOGI

1. Sumber Data

Data yang digunakan dalam artikel ini adalah data historis nasabah terkait catatan biro kredit dan hasil pinjaman suatu bank, yaitu bank "ABC". Data ini diperoleh dari dataset yang digunakan dalam channel youtube Skillcate dengan video yang berjudul "Credit Scoring Project using Machine Learning".

2. Variabel

Terdapat 30 variabel dalam dataset yang digunakan. Variabel ini dituliskan dalam Tabel 1.

Tabel 1 Variabel

Variable Name	Label	Role	Variable Name	Label	Role
Target	Target = 1 (Defaulters), Target = 0 (Good Loans)	Target	TLSatCnt	Number Trade Lines Currently Satisfactory	Input
BanruptcyInd	Bankruptcy Indicator	Input	TLCnt12	Number Trade Lines Opened 12 Months	Input
TLBadDerogCnt	Bad Dept plus Public Derogatories	Input	TLCnt24	Number Trade Unes Opened 24 Months	Input
CollectCnt	Collections	Input	TLCnt03	Number Trade Unes Opened 3 Months	Input
InqFinanceCnt24	Finance Inquires 24 Months	Input	TSatPct	Percent Satisfactory to Total Trade Lines	Input
InqCnt06	Inquiries 6 Months	Input	TLBalHCPct	Percent Trade Line Balance to High Credit	Input
DerogCnt	Number Public Derogatories	Input	TLOpenPct	Percent Trade Lines Open	Input
TLDe13060Cnt24	Number Trade Lines 30 or 60 Days 24 Months	Input	TLOpen24Pct	Percent Trade Unes Open 24 Months	Input
TL50Utilent	Number Trade Lines 50 pct Utilized	Input	TLTimeFirst	Time Since First Trade Une	Input
TDe160Cnt24	Number Trade Lines 60 Days or Worse 24 Months	Input	InqTimeLast	Time Since Last Inquiry	Input
TDe160CntAll	Number Trade Lines 60 Days or Worse Ever	Input	TLTimeLast	Time Since Last Trade Line	Input
T75UtilCnt	Number Trade Unes 75 pct Utilized	Input	TLSum	Total Balance All Trade Lines	Input
Tel90Cnt24	Number Trade Unes 90+ 24 Months	Input	TLMaxSum	Total High Credit All Trade Lines	Input
TLBadCnt24	Number Trade Unes Bad Debt 24 Months	Input	TL_Cnt	Total Open Trade Unes	Input
TLDe160Cnt	Number Trade Unes Currently 60 Days or Worse	Input	ID	Customer ID	ID

Variabel yang digunakan terbagi menjadi dua, yaitu variabel respon dan variabel prediktor. Variabel respon adalah variabel dengan *role* sebagai 'Target', sedangkan variabel prediktor adalah variabel dengan *role* sebagai 'Input'. Adapun variabel ID tidak dimasukkan dalam pemodelan karena variabel ini tidak memiliki pengaruh terhadap variabel respon. Berikut merupakan penjelasan dari variabel-variabel tersebut.

- Bankruptcy indicator*, yaitu angka yang menunjukkan kemungkinan peminjam mengajukan kebangkrutan.
- Public derogatories*, yaitu jumlah keterlambatan pembayaran atau *default*.

- c. *Financial enquires*, yaitu jumlah permintaan kredit yang dibuat dalam beberapa bulan terakhir.
- d. *Trade line*, yaitu angka yang menunjukkan jumlah rekening kredit yang dimiliki peminjam, baik itu pinjaman, kartu kredit, atau kewajiban hutang lainnya.
- e. *Target*, yaitu kategori hasil pinjaman (0='good loan' dan 1='bad loan')

3. Langkah Penyelesaian

Untuk menyelesaikan permasalahan di bank "ABC", perlu dilakukan beberapa langkah sebagai berikut.

- 1) Merumuskan masalah dari sudut pandang bisnis maupun dari sudut pandang metodologi (variabel, metode, dll).
- 2) Mengumpulkan data.
- 3) Melakukan persiapan data (*data preparation*), yaitu membuang variabel ID, mengecek *missing values*, dan mengatasi adanya *missing values* dengan mengganti nilai tersebut dengan mean dari masing-masing variabel terkait.
- 4) Melakukan eksplorasi data.
- 5) Melakukan modeling *machine learning* sebagai berikut.
 - a. Membuat desain *training set* dan *testing set* dari jumlah data yang terkumpul dengan perbandingan 80:20.
 - b. Melakukan standarisasi data dengan transformasi.
 - c. Membuat model klasifikasi dan prediksi menggunakan regresi logistik.
 - d. Melakukan evaluasi model menggunakan *confusion matrix* untuk mengecek hasil klasifikasi dan menggunakan nilai akurasi untuk mengetahui persentase ketepatan klasifikasi.
 - e. Menarik kesimpulan dari model yang terbentuk.
- 6) Melakukan *prescriptive analytics* dengan *decile methodology*.
- 7) Mempresentasikan hasil dan solusi atas permasalahan yang dihadapi.

KESIMPULAN

Data science merupakan suatu ilmu yang sangat berguna untuk mengolah data, terutama *big data*, agar menjadi informasi yang berguna dalam pengambilan keputusan. *Data science* banyak digunakan di berbagai industri. Dalam industri perbankan, *data science* digunakan salah satunya dalam *risk management*, yaitu *credit scoring*. *Credit scoring* adalah suatu nilai yang dapat digunakan perusahaan untuk menentukan nasabah mana yang berhak mendapatkan pinjaman. Dengan *credit scoring*, perusahaan juga dapat menentukan perubahan suku bunga atau batas kredit pada kartu kredit nasabah.

DAFTAR PUSTAKA

- [1] D. Kurniasari, "Pengertian Data Science, Contoh Penerapannya, Dan Bidang Profesi Terkait Data Science," DQLab, 31 March 2021. [Online]. Available: <https://dqlab.id/pengertian-data-science-contoh-dan-profesi>. [Diakses 19 March 2023].

- [2] S. M. Rezkia, “Tujuan Utama dari Data Science dengan Kenali Keunggulannya dalam Industri Perbankan,” DQLab, 6 August 2021. [Online]. Available: <https://dqlab.id/tujuan-utama-dari-data-science-dengan-kenali-keunggulannya-dalam-industri-perbankan>. [Diakses 19 March 2023].
- [3] The Investopedia Team, “Credit Score: Definition, Factors, and Improving It,” Investopedia, 12 September 2022. [Online]. Available: https://www.investopedia.com/terms/c/credit_score.asp. [Diakses 19 March 2023].
- [4] Skillcate, Penyusun, *Credit Scoring Project using Machine Learning | Risk Modelling | Logistic Regression | ML Education*. [Perekaman Suara]. Skillcate. 2021.
- [5] Dicoding Intern, “Apa itu Machine Learning? Beserta Pengertian dan Cara Kerjanya,” Dicoding, 19 August 2020. [Online]. Available: <https://www.dicoding.com/blog/machine-learning-adalah/>. [Diakses 19 March 2023].
- [6] N. Rahmalia, “Machine Learning: Definisi, Bedanya dengan AI, dan Tips Mempelajarinya,” Glints, 13 February 2023. [Online]. Available: <https://glints.com/id/lowongan/machine-learning/#.ZBcb1nZBy01>. [Diakses 19 March 2023].
- [7] DQLab Team, “Rekomendasi Ragam Algoritma Machine Learning Terupdate,” DQLab, 13 September 2022. [Online]. Available: <https://dqlab.id/rekomendasi-ragam-algoritma-machine-learning-terupdate>. [Diakses 19 March 2023].