

INITIAL ASSESSMENT

Data Analytics_Ardia Fatma Sari

ESAI

USE CASE CREDIT SCORING DALAM MENUNJANG BISNIS

Credit scoring (penilaian kredit) adalah bagian dari lanskap keuangan pada zaman sekarang dan digunakan dalam dunia industri serta departemen organisasi. Credit scoring digunakan untuk mengevaluasi risiko kredit dalam hal permohonan pinjaman dari konsumen (Brondel, n.d.). Banyak sekali perusahaan keuangan, ritel, dan sebagainya memanfaatkan *credit scoring* untuk memprediksi segala sesuatu yang diperlukan atau dimungkinkan terjadi dalam suatu perusahaan. Dengan adanya prediksi dapat membantu perusahaan dalam pengambilan setiap keputusan.

Credit scoring mengelompokkan keragaman dari karakteristik konsumen yang mengajukan permintaan kredit berdasarkan kesalahan dari kelalaian atas kewajiban. Credit scoring dibentuk melalui serangkaian proses statistika dimana itu digunakan untuk melakukan prediksi terhadap data baru. Dalam membuat credit scoring business dibutuhkan data analisis untuk merancang dan mengembangkan model yang akurat dan bermanfaat. Dalam pembuatannya, data analisis harus bisa memastikan model yang didapatkan itu mudah dinilai dan masuk akal untuk digunakan. Dalam penentuan model, dibutuhkan machine learning. Tujuan khusus model credit scoring untuk melakukan klasifikasi, hal itu digunakan untuk membedakan calon peminjam yang baik dan buruk. Model statistik diperlukan dalam menentukan garis pemisah yang membedakan dua kategori tersebut dalam ruang variabel penjelas. (Saputro, 2011)

Dalam melakukan pengembangan credit scoring, ada beberapa langkah-langkah yang harus dilakukan, antara lain sebagai berikut. **Pertama** adalah Pemahaman Bisnis. Dalam credit scoring diperlukan untuk memahami kebutuhan bisnis yang dilayani oleh credit scoring engine. Pemahaman ini termasuk tujuan bisnis, penggunaan credit scoring engine, dan kriteria dari kredit yang diinginkan. **Kedua**, Pemahaman data. Dalam tahap ini perlu ada pemilihan data. Pemilihan data adalah kegiatan dalam menentukan data set yang akan digunakan untuk memperoleh model sebagai bahan dalam memprediksi sesuatu. Setelah menentukan data set yang perlu dilakukan adalah cleaning data terlebih dahulu kemudian transformasi data. Dalam tahap ini, ada beberapa hal yang harus dilakukan, antara lain yaitu perlu melakukan check missing value. Ada dua cara untuk melakukannya, antara lain adalah take out variabel (membuang variabel yang memiliki jumlah missing value yang cukup banyak) dan complete case (membuang observasi yang missing value). Kemudian, menyesuaikan tipe data, variabel yang dimiliki harus disesuaikan satu sama lain. Selanjutnya adalah exploratory data. Dalam tahap ini dilakukan visualisasi data baik untuk data kategorik maupun numerik, kemudian dilakukan cross validation. Data yang digunakan tadi dibagi menjadi data training dan data testing (Data Science, 2021). **Ketiga** adalah tahap modelling. Setelah data siap, pengembangan dilakukan untuk memprediksi risiko kredit. Teknik yang dapat digunakan untuk mengembangkan model antara lain adalah decision tree, regresi logistik, algoritma random forest, classification, neural network, dan lain sebagainya. Saat melakukan tahap modelling, model yang didapatkan tadi bisa dicompare. Tahap **keempat** adalah evaluasi. Beberapa model yang sudah dicompare dibandingkan untuk melihat model yang terbaik. Kemudian dilakukan validitas model, membuat strategi risiko, dan analisis ROI (Return of Investment). Tahap **Kelima** adalah implementasi dan pengembangan sistem. Setelah pengujian model dan didapatkan model terbaik, selanjutnya adalah pengimplementasian model. Dalam hal ini, perlu

melibatkan integrasi model ke sistem dan mengembangkan untuk pengguna (ALTAIR, 2022). Untuk memastikan fungsinya, maka perlu adanya pengujian sistem. Selanjutnya adalah langkah **keenam** yaitu dashboard monitoring. Tujuannya adalah untuk melacak kinerja sistem dan memonitor key performance indicator. Dashboard monitoring dapat memvisualisasikan perkembangan model secara realtime. Pemantauan dashboard secara berkala dapat membantu dalam pengambilan keputusan dan mengoptimalkan parameter dalam model agar lebih akurat.

Dalam business credit scoring engine development dibutuhkan untuk memahami bisnis, memahami data yang digunakan dan mempersiapkannya dengan baik, menentukan model yang terbaik kemudian mengimplementasikannya ke sistem dan mengembangkan dashboard monitoring. Dengan melakukan semua langkah-langkah yang dibutuhkan, maka akan dihasilkan suatu sistem yang efektif untuk memprediksi risiko dan memantau performa model.

DAFTAR PUSTAKA

ALTAIR. (2022, May 18). *Credit Scoring Series Part Two: Credit Scorecard Modeling*

Methodology. <https://www.altair.com/newsroom/articles/credit-scoring-series-part-two-credit-scorecard-modeling-methodology>

Brondel, P. (n.d.). *Credit scoring—Case study in data analytics*.

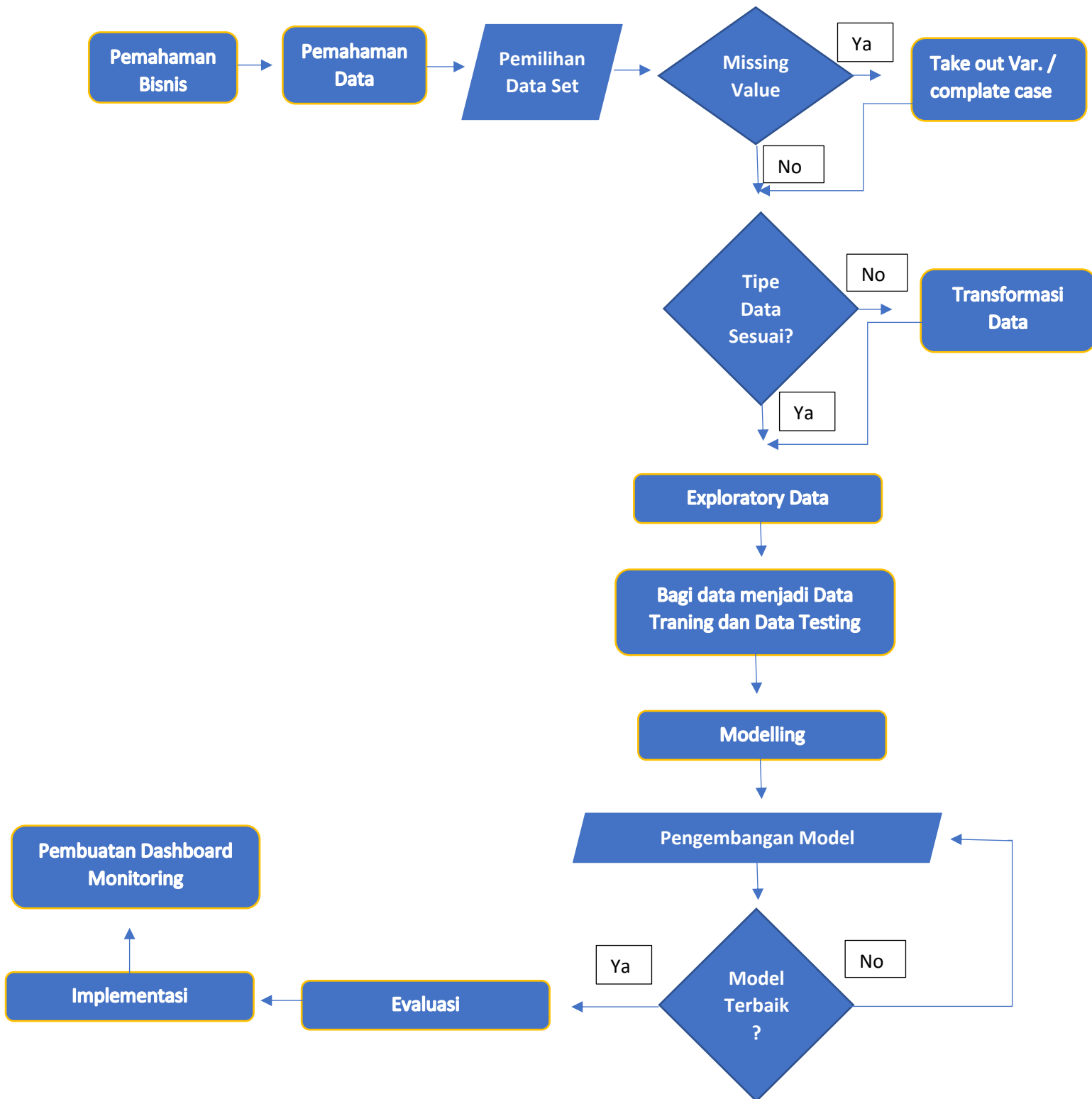
Data Science, A. (Director). (2021). *CREDIT SCORING / Algoritma Show / Episode #1 /*

Algoritma 2021. <https://youtu.be/L7564DMRcdY>

Ramaquita, A., & Alamsyah, A. (n.d.). *MODEL CREDIT SCORING BERDASARKAN DATA DEMOGRAFI DAN JEJARING SOSIAL DI MEDIA SOSIAL (STUDI KASUS: LINKEDIN)*.

Saputro, D. (2011). *MODEL CREDIT SCORING UNTUK PROSES ANALISA KELAYAKAN FASILITAS KREDIT MOTOR MENGGUNAKAN METODE CLASSIFICATION AND REGRESSION TREE (CART) (Studi Kasus PT. X Finance Mauk Desember 2010)*.

DIAGRAM ALIR BUSINESS CREDIT SCORING



Contoh Output Model Credit Scoring Berdasarkan Data Media Sosial

(Ramaquita & Alamsyah, n.d.)

Bisnis yang dilakukan ini memerlukan media sosial sebagai sarana promosi. Media social yang digunakan adalah LinkedIn. Data yang digunakan adalah data demografi dan jejaring social yang memiliki insight karakter, kapasitas, dan kondisi.

No	Atribut	Nilai	Keterangan	No.	Atribut	Nilai	Keterangan
1	Gender	1	Wanita	5	Degree Centrality	1	0 - 5
		2	Pria			2	5 - 10
2	Age	3				3	> 10
		1	< 19	6	Betweenness Centrality	1	0 - 1
		2	19 - 34			2	1 - 10
		3	35 - 54			3	> 10
		4	> 54	7	Closeness Centrality	1	0 - 0.51
3	Job	1	0 - 5.000.000			2	0.51 - 0.53
		2	5.000.0001 - 10.000.000			3	> 0.53
		3	10.000.001 - 20.000.000				
		4	20.000.001 - 40.000.000				
		5	> 40.000.000				
4	Education		*dalam rupiah	8	Clustering Coefficient	1	0 - 0.5
		1	Khuster 5			2	0.5 - 0.7
		2	Khuster 4			3	> 0.7
		3	Khuster 3				
		4	Khuster 2				
		5	Khuster 1				

Setelah data didapatkan, dilakukan pembersihan data yang tidak relevan dan kosong. Kemudian data di transformasi jika diperlukan. Kemudian dilakukan pembentukan fitur data. Dalam kasus ini, dibentuklah fitur data menjadi 3 jenis data input model sebagai berikut.

1. Fitur demografi
2. Fitur jejaring social
3. Fitur demografi dan jejaring social

Setelah itu dilakukan komparasi model dengan menggunakan algoritma decision tree dan random forest.

Algoritma	Model Percobaan	Performance Matrix			
		Accuracy	Precision	Recall	F-Measure
Decision Tree	Model 1	70.45	65.38	61.82	63.55
	Model 2	61.36	55.00	40.00	46.32
	Model 3	73.48	71.74	60.00	65.35
Random Forest	Model 1	68.18	61.82	61.82	61.82
	Model 2	65.91	59.26	58.18	58.72
	Model 3	77.27	76.60	65.45	70.59

Didapatkan hasil nilai evaluasi kinerja *performance metrics* tertinggi. Dari hasil pengolahan, model 3 menghasilkan nilai *accuracy* sebesar 77.27 persen, nilai *precision* 76.60

persen, nilai *recall* sebesar 65.45 persen, dan nilai *f-measure* sebesar 70.59 persen. Selain itu, dari hasil komparasi model 1 dan 2 juga dapat disimpulkan bahwa fitur jejaring sosial pada model 2 cenderung lebih baik apabila dijadikan data pelengkap untuk meningkatkan hasil nilai evaluasi kinerja mengingat hasil pada model 2 memiliki tingkat *accuracy* data yang tidak terlampau jauh dengan model 1.