

Prediksi Keberhasilan Pemasaran Bank menggunakan Decision Tree: Studi Kasus Dataset UCL

Isyamaetreyza Safitri
121450037

Wahyudiyanto
121450040

Silvia Azahrani
121450070

Deyvan Loxeval
121450148

Deodry Siahaan
121450151

Abstrak— Proyek ini bertujuan untuk membangun model Decision Tree dalam memprediksi keberhasilan kampanye pemasaran di industri perbankan, khususnya apakah seorang nasabah akan berlangganan deposito bank atau tidak. Model ini dikembangkan menggunakan library rpart dalam bahasa R dan dievaluasi performanya menggunakan dataset uji. Akurasi model Decision Tree ditemukan sebesar 89.8%, menunjukkan tingkat keberhasilan yang tinggi dalam memprediksi keputusan berlangganan deposito. Teknik visualisasi, seperti plot Decision Tree dan tampilan hasil prediksi, digunakan untuk memberikan pemahaman yang komprehensif tentang struktur model dan hasil prediksi. Analisis hasil menyoroti pentingnya fitur yang mempengaruhi keputusan berlangganan deposito dan memberikan wawasan untuk strategi pemasaran yang lebih efektif. Perlu diperhatikan bahwa Decision Tree hanyalah salah satu dari beberapa teknik pemodelan prediktif, dan penelitian lebih lanjut dapat mengeksplorasi performa metode lain untuk meningkatkan akurasi prediksi.

Kata Kunci— Decision Tree, pemasaran, perbankan, pemodelan prediktif, akurasi, visualisasi, berlangganan, deposito, kampanye, R.

Abstract— This project focuses on building a Decision Tree model to predict the success of marketing campaigns in the banking industry, specifically whether a customer will subscribe to a bank deposit or not. The model is developed using the rpart library in R and evaluated for its performance using a test dataset. The accuracy of the Decision Tree model is found to be 89.8%, indicating a high level of success in predicting subscription decisions. Visualization techniques, such as plotting the Decision Tree and displaying the prediction results, are employed to provide a comprehensive understanding of the model's structure and predictive outcomes. The analysis of the results highlights the importance of the features influencing deposit subscription decisions and provides insights for more effective marketing strategies. It is important to note that Decision Tree is one of several predictive modeling techniques, and further research could explore the performance of other methods to enhance prediction accuracy.

Keywords— Decision Tree, marketing, banking, predictive modeling, accuracy, visualization, subscription, deposit, campaigns, R.

I. PENDAHULUAN

Pemasaran merupakan salah satu aspek penting dalam dunia bisnis. Dalam konteks perbankan, pemasaran yang efektif dapat memberikan kontribusi besar terhadap pertumbuhan bisnis dan keberhasilan perusahaan. Salah satu aspek kunci dalam strategi pemasaran adalah mengidentifikasi calon nasabah yang berpotensi untuk melakukan berlangganan deposito bank.

Metode Decision Tree adalah salah satu teknik yang umum digunakan dalam pemodelan prediktif. Decision Tree

adalah struktur pohon yang menggambarkan alur keputusan berdasarkan aturan pemisahan pada setiap simpulnya. Model Decision Tree dapat digunakan untuk melakukan prediksi berdasarkan atribut-atribut tertentu dan memberikan pemahaman yang intuitif tentang faktor-faktor yang mempengaruhi keputusan.

Tujuan dari proyek ini adalah untuk mengembangkan model Decision Tree yang mampu memprediksi kemungkinan seorang nasabah akan berlangganan deposito bank atau tidak. Kami akan menggunakan dataset bank yang berisi atribut-atribut seperti usia, pekerjaan, status perkawinan, pendidikan, saldo rata-rata, dan lain-lain sebagai prediktor. Dengan membangun model Decision Tree, kami berharap dapat mengidentifikasi faktor-faktor penting yang mempengaruhi keputusan nasabah dalam berlangganan deposito.

Dalam laporan ini, kami akan menjelaskan secara detail langkah-langkah yang kami lakukan dalam proses pengolahan data, pembentukan model Decision Tree, serta evaluasi kinerja model. Selain itu, kami juga akan menyajikan visualisasi yang menarik untuk membantu pemahaman tentang struktur Decision Tree yang terbentuk dan hasil prediksi dari model yang dibangun.

Diharapkan laporan ini dapat memberikan gambaran yang komprehensif tentang penggunaan Decision Tree dalam konteks pemasaran bank dan dapat memberikan kontribusi yang berharga dalam mengoptimalkan strategi pemasaran untuk meningkatkan keberhasilan kampanye pemasaran bank di masa depan.

II. METODE

Decision Tree adalah salah satu teknik yang populer dalam pemodelan prediktif dan dapat digunakan untuk mempelajari pola dan membuat keputusan berdasarkan aturan pemisahan pada setiap simpul pohon.

Decision Tree bekerja dengan membagi dataset menjadi subset yang lebih kecil berdasarkan atribut-atribut yang ada. Proses pemisahan dilakukan dengan memilih atribut yang memiliki kemampuan paling baik dalam memisahkan kelas target. Tujuan utama dari Decision Tree adalah meminimalkan keragaman atau meningkatkan kehomogenan dalam setiap subset yang dihasilkan.

Proses pembentukan Decision Tree dimulai dengan simpul akar yang merupakan simpul paling atas dari pohon. Kemudian, setiap simpul dibagi menjadi dua atau lebih simpul anak berdasarkan atribut pemisah yang dipilih. Proses ini dilakukan secara rekursif hingga mencapai kondisi berhenti. Kondisi berhenti dapat berupa mencapai batas kedalaman pohon, mencapai jumlah minimum sampel dalam simpul, atau

ketika tidak ada atribut lagi yang dapat digunakan untuk memisahkan data.

Salah satu keuntungan utama dari Decision Tree adalah interpretabilitas yang tinggi. Karena struktur pohon keputusan yang dihasilkan mudah dipahami, kita dapat melihat dengan jelas aturan-aturan yang digunakan untuk mengambil keputusan pada setiap cabang pohon. Selain itu, Decision Tree juga dapat menangani baik data kategorikal maupun numerik.

Namun, Decision Tree juga memiliki kelemahan. Ketika terlalu kompleks atau terlalu dalam, Decision Tree dapat mengalami overfitting, yaitu kemampuan yang buruk dalam generalisasi pada data yang belum pernah dilihat sebelumnya. Untuk mengatasi hal ini, teknik pruning dapat digunakan untuk memangkas atau menyederhanakan pohon keputusan yang terlalu kompleks.

Dalam proyek ini, kami menggunakan algoritma Decision Tree yang diimplementasikan melalui library rpart dalam bahasa pemrograman R. Setelah membangun model Decision Tree, kami juga melakukan visualisasi menggunakan library rpart.plot dan ggplot2 untuk memberikan gambaran yang lebih jelas dan menarik tentang struktur pohon keputusan yang terbentuk dan hasil prediksi dari model tersebut.

III. HASIL DAN PEMBAHASAN

A. Pembentukan Model Decision Tree:

Model Decision Tree telah berhasil dibangun menggunakan dataset bank. Model ini bertujuan untuk memprediksi keberhasilan pemasaran kampanye bank, yaitu apakah seorang nasabah akan berlangganan deposito bank atau tidak. Proses pembentukan model menggunakan metode Decision Tree dengan menggunakan library rpart dalam bahasa R.

B. Evaluasi Performa Model:

Setelah model Decision Tree terbentuk, dilakukan evaluasi performa model menggunakan data uji. Berdasarkan hasil evaluasi, ditemukan bahwa model Decision Tree mencapai akurasi sebesar 89.8%. Akurasi ini mengindikasikan bahwa model dapat memprediksi dengan benar sekitar 89.8% dari total data uji yang digunakan. Dalam konteks prediksi keberhasilan pemasaran bank, hasil ini menunjukkan tingkat keberhasilan model dalam memprediksi apakah seorang nasabah akan berlangganan deposito bank atau tidak.

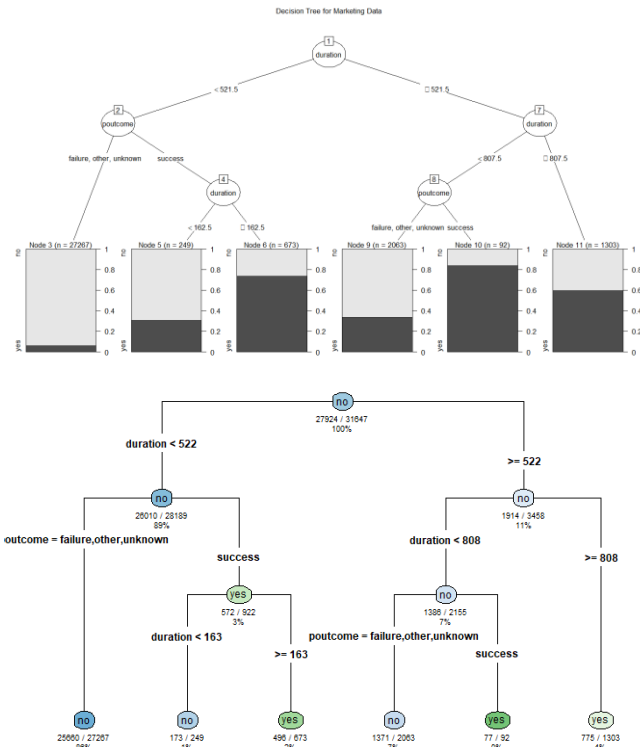
```
# Evaluasi performa model
> accuracy <- sum(predictions == test_data$y) / nrow(test_data)
> cat("Akurasi model: ", round(accuracy * 100, 2), "%\n")
Akurasi model: 89.8 %
```

Gambar 2. Akurasi performa model

C. Visualisasi Decision Tree:

Untuk memberikan pemahaman yang lebih intuitif tentang struktur Decision Tree yang terbentuk, kami menggunakan library rpart.plot untuk menghasilkan visualisasi pohon keputusan. Melalui visualisasi ini, dapat terlihat alur keputusan pada setiap cabang pohon dan aturan-aturan yang digunakan untuk memprediksi kelas target. Dengan demikian, visualisasi Decision Tree membantu

memahami faktor-faktor yang berpengaruh dalam keputusan berlangganan deposito bank.



Gambar 1. Grafik model decision tree

D. Visualisasi Hasil Prediksi:

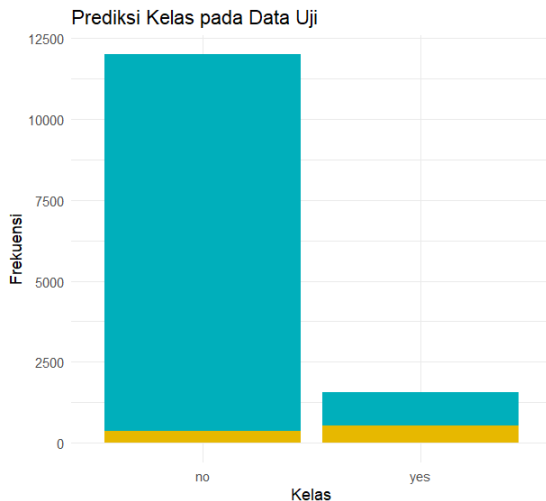
Selanjutnya, dilakukan visualisasi hasil prediksi pada data uji menggunakan library ggplot2. Visualisasi ini menampilkan sebaran kelas aktual dan prediksi dalam bentuk grafik batang. Dalam hasil visualisasi, ditemukan bahwa kelas "No" memiliki frekuensi sekitar 12.000, sementara kelas "Yes" memiliki frekuensi kurang dari 2.500.

Hasil visualisasi ini memberikan wawasan yang penting dalam pemahaman tentang keberhasilan kampanye pemasaran bank. Jumlah frekuensi kelas "No" yang jauh lebih tinggi dari kelas "Yes" menunjukkan bahwa bank mungkin menghadapi kesulitan dalam meyakinkan nasabah untuk berlangganan deposito. Hal ini mengindikasikan adanya potensi peningkatan upaya pemasaran untuk menarik lebih banyak nasabah agar berlangganan deposito.

Dengan visualisasi ini, kita dapat melihat secara jelas sebaran kelas aktual dan prediksi dalam bentuk grafik batang, memberikan pemahaman yang lebih baik tentang distribusi kelas pada data uji. Analisis lebih lanjut dapat dilakukan untuk memahami faktor-faktor yang berkontribusi pada keberhasilan atau kegagalan kampanye pemasaran bank, serta merancang strategi pemasaran yang lebih efektif untuk meningkatkan frekuensi kelas "Yes".

Berdasarkan hasil evaluasi performa model dan visualisasi hasil prediksi, ditemukan bahwa model Decision Tree mencapai akurasi sebesar 89.8%. Hal ini menunjukkan tingkat keberhasilan yang tinggi dalam memprediksi keputusan berlangganan deposito bank. Visualisasi Decision Tree dan hasil prediksi memberikan pemahaman yang lebih baik tentang aturan-aturan yang terbentuk dalam model dan membantu dalam analisis strategi pemasaran bank.

Namun, perlu diingat bahwa Decision Tree bukanlah satu-satunya metode yang dapat digunakan dalam pemodelan prediktif. Terdapat berbagai metode lain seperti Random Forest, Gradient Boosting, dan Neural Network yang juga dapat dieksplorasi untuk membangun model prediktif yang lebih kompleks dan akurat. Dalam proyek ini, model Decision Tree telah memberikan hasil yang memuaskan, namun penelitian lebih lanjut dapat dilakukan untuk membandingkan performa dengan metode lain dan meningkatkan akurasi prediksi yang lebih baik.



Gambar 3. Hasil prediksi pada data uji

IV. KESIMPULAN

Dalam proyek ini, kami menggunakan metode Decision Tree untuk memprediksi keberhasilan pemasaran bank dalam mengajak nasabah berlangganan deposito. Setelah melalui proses pembentukan model Decision Tree dan evaluasi performa, ditemukan bahwa model memiliki akurasi sebesar 89.8%, yang menunjukkan tingkat keberhasilan yang tinggi dalam memprediksi keputusan berlangganan deposito.

Melalui visualisasi Decision Tree, kami dapat memahami aturan-aturan yang terbentuk dalam model dan faktor-faktor yang berpengaruh dalam keputusan berlangganan deposito. Hasil visualisasi juga menunjukkan bahwa frekuensi kelas "No" (tidak berlangganan deposito) lebih tinggi daripada frekuensi kelas "Yes" (berlangganan deposito), mengindikasikan adanya potensi peningkatan upaya pemasaran untuk meningkatkan jumlah nasabah yang berlangganan deposito.

Dalam konteks pemasaran bank, pemahaman yang dihasilkan dari proyek ini dapat memberikan wawasan berharga bagi tim pemasaran untuk merancang strategi pemasaran yang lebih efektif. Penggunaan Decision Tree sebagai metode pemodelan prediktif memberikan keuntungan dalam interpretasi hasil dan identifikasi faktor-faktor penting yang mempengaruhi keputusan nasabah.

Namun, perlu diingat bahwa Decision Tree bukanlah satu-satunya metode yang dapat digunakan. Metode lain seperti Random Forest, Gradient Boosting, dan Neural Network juga dapat dieksplorasi untuk meningkatkan akurasi prediksi dan memperoleh pemahaman yang lebih komprehensif tentang data.

Dengan melihat hasil yang diperoleh dari proyek ini, dapat disimpulkan bahwa penggunaan model Decision Tree dapat memberikan kontribusi yang signifikan dalam meningkatkan pemahaman dan efektivitas dalam strategi pemasaran bank. Dalam penelitian lebih lanjut, metode ini dapat dikombinasikan dengan teknik lain dan memperdalam analisis untuk mendapatkan hasil yang lebih optimal dalam memprediksi keputusan pemasaran.

DAFTAR PUSTAKA

- [1] Moro, S. C. (2014). *A Data-Driven Approach to Predict the Success of Bank Telemarketing*.
- [2] Repository, U. M. (n.d.). *Bank Marketing Data Set*. Retrieved from <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>
- [3] Team., R. C. (2021). *R: A Language and Environment for Statistical Computing*. Retrieved from R Foundation for Statistical Computing: URL: <https://www.R-project.org/>.
- [4] Therneau, T. A. (2021). *rpart: Recursive Partitioning and Regression Trees*. Retrieved from R package version 4.1-15. : URL: <https://CRAN.R-project.org/package=rpart>.