



Prediksi Peningkatan Jumlah Nasabah Deposito Berjangka Menggunakan Algoritma KNN, *Decision Tree*, *Random Forest* Dan *Xgboost*

Nurlaelatul Maulidah^{1*}

¹Sistem Informasi Akuntansi, Universitas Bina Sarana Informatika Kampus Kota Tegal
Jl. Sipelem No.22, Kraton, Kec. Tegal Barat, Kota Tegal, Jawa Tengah 52112

*nurlaelatulmaulidah1@gmail.com

Abstrak:

Bank merupakan sebuah lembaga keuangan yang umumnya didirikan untuk menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkan kepada masyarakat dalam bentuk kredit atau bentuk lainnya dengan rangka meningkatkan taraf hidup rakyat banyak. Pada penelitian ini, dilakukan pengujian empat algoritma *machine learning* yaitu *K-Nearest Neighbor* (K-NN), *Decision Tree*, *Random Forest* dan *XGBoost*, untuk mengetahui dan membandingkan tingkat akurasi dari masing-masing algoritma tersebut dalam melakukan prediksi terhadap peningkatan jumlah nasabah deposito berjangka bank. Pada penelitian ini *dataset* diperoleh dari *UCI Machine Learning Repository*. Data yang diperoleh kemudian diproses. Dari hasil pengujian didapatkan tingkat akurasi terbaik sebesar 92,36% dengan menggunakan algoritma *XGBoost*.

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license



Keywords:

Deposito Berjangka;
KNN;
Decision Tree;
Random Forest;
XGBoost;

Article history:

Diserahkan 30 Agustus 2022
Direvisi 06 Juni 2023
Diterima 12 Juni 2023
Dipublikasi 15 Agustus 2023

DOI:

10.22441/incomtech.v13i3.16921

1. PENDAHULUAN

Bank adalah badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan, dan menyalurkannya kepada masyarakat dalam rangka meningkatkan taraf hidup masyarakat [1]. Bank sebagai lembaga keuangan akan selalu bersaing untuk mendapatkan kepercayaan dari masyarakat dan pengusaha pemilik modal agar menyalurkan dananya kepada pihak yang memerlukan. Masyarakat maupun perusahaan sebagai pemilik dana mempunyai suatu keinginan, agar dana yang ada dapat berkembang. Bertambahnya nilai suatu dana merupakan suatu perkembangan yang diinginkan oleh para pemilik dana baik dalam jangka pendek maupun jangka panjang.

Di tengah pesatnya perkembangan penghimpunan dana dan penyaluran kredit perbankan, pemerintah mengeluarkan kebijaksanaan baru, yang bertujuan untuk melengkapi dan menyempurnakan perkembangan dalam sektor perbankan.

Nasabah lebih memilih menginvestasikan kelebihan dananya pada tempat yang memberikan keuntungan yang besar di sertai dengan rasa aman. Perbankan merupakan salah satu tempat populer menurut UU RI No 10 Tahun 1988 tanggal 10 November 1988 tentang perbankan, dapat disimpulkan bahwa usaha perbankan meliputi tiga kegiatan, yaitu menghimpun dana, menyalurkan dana, dan memberikan jasa bank lainnya.

Kegiatan menghimpun dan menyalurkan dana bisa diasumsikan sebagai kegiatan pokok bank sedangkan memberikan jasa bank lainnya hanya kegiatan pendukung. Kegiatan penghimpun dana, berupa mengumpulkan dana dari masyarakat dalam bentuk simpanan giro, tabungan, dan deposito. Bank salah satu sarana yang memudahkan aktivitas masyarakat untuk menyimpan uang, dalam hal perniagaan maupun untuk investasi masa depan. Dalam upaya menarik minat nasabah untuk berlangganan deposito berjangka di bank dilakukan berbagai upaya. Salah satunya yaitu pendekatan menggunakan Teknik telemarketing.

Deposito adalah simpanan pihak ketiga pada bank yang penarikannya hanya dapat dilakukan dalam jangka waktu tertentu menurut perjanjian antara penyimpan (pihak ketiga) dengan bank yang bersangkutan [1]. Jadi, Deposito berjangka merupakan salah satu produk perbankan yang memiliki jangka waktu untuk pengambilan dana. Jangka waktu penarikannya sesuai kesepakatan yang telah ditetapkan. Nasabah bebas menentukan waktu penarikannya sesuai keinginan atau kebutuhan mereka, mulai dari 1, 3, 6, dan 12 sampai 24 bulan.

Hal inilah yang menjadi landasan masalah bagi penulis dalam melakukan penelitian yaitu mengenai Prediksi Jumlah Nasabah Deposito Berjangka Menggunakan Algoritma *K-Nearest Neighbor* (K-NN), *Decision Tree*, *Random Forest* Dan *XGBoost*.

2. METODE

a. *K-Nearest Neighbor* (KNN)

K-Nearest Neighbor (K-NN) adalah suatu metode yang menggunakan algoritma supervised dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN [2].

Sedangkan, Menurut Astuti dan Guntara dalam Isman, Ahmad dan Latief “*K-Nearest Neighbor* adalah sebuah metode untuk melakukan Klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran di proyeksikan keruang berdimensi banyak, dimana masing masing dimensi merepresentasikan fitur dari data [3]”.

b. *Decision Tree*

Menurut Arni dalam Sandag, “Algoritma *Decision Tree* adalah metode pembelajaran mesin dalam membuat keputusan menggunakan model seperti pohon. *Decision Tree* juga bisa menyelesaikan berbagai alternatif mengatasi masalah, juga terdapat faktor-faktor yang mempengaruhi perhitungan [4]”. Jarak dalam *Decision Tree* dihitung menggunakan Entrophy dengan rumus:

$$Entropy (S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (1)$$

Keterangan:

S : Himpunan kasus

k : Jumlah partisi S

Pj : Probabilitas yang didapat dari jumlah (Ya/Tidak) dibagi total kasus

c. *Random Forest*

Random Forest adalah salah satu algoritma machine learning yang sangat baik dalam memecahkan masalah klasifikasi dan regresi [5].

Sedangkan, Menurut Saleh dalam Irawan, et al, “Metode *random forest* merupakan pengembangan dari decision tree dimana setiap *decision tree* telah dilakukan proses pelatihan dengan menggunakan sampel individu. *Random forest* yang dihasilkan memiliki banyak tree dan setiap tree ditanam dengan cara yang sama. Seiring dengan bertambahnya dataset, maka *tree* juga ikut berkembang [6]”. Sedangkan Cutler dan Stevens mengemukakan bahwa *Random Forest* adalah kumpulan dari pohon klasifikasi hasil dari *sampling bootstrap data*. Langkah awal dalam membangun model *random forest* yaitu menentukan nilai N sebagai jumlah *decision tree* yang dibangun. Pemilihan data yang digunakan untuk pembangunan *tree* menggunakan teknik *bootstrap sample* Teknik ini akan memilih sampel dari data secara acak dan dilakukan secara berulang hingga jumlah sampel pada *bootstrap sample* sama dengan jumlah data sebenarnya. Tujuan pemilihan ini bukan untuk mencari kemungkinan terbaik dari split semua data, namun hanya mencari kemungkinan terbaik dari split dari gen-gen yang terpilih [7].

d. *XGBoost*

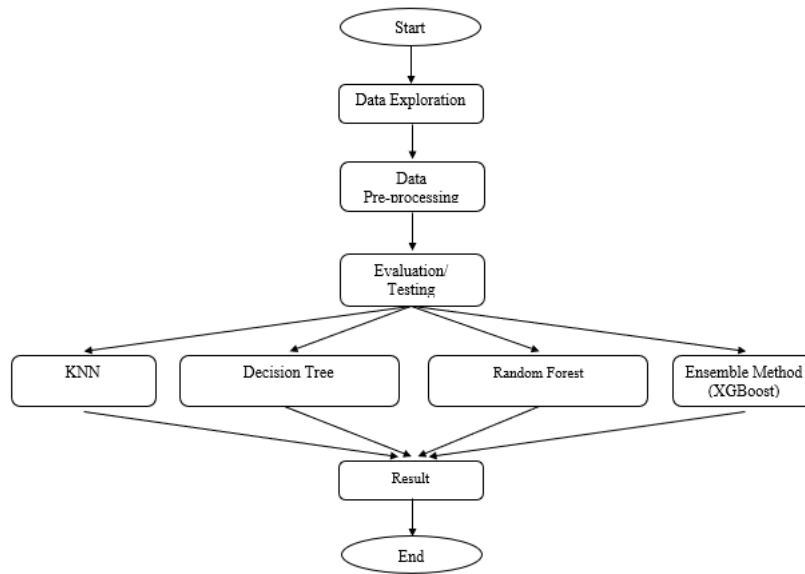
XGboost merupakan salah satu metode *boosting* yaitu kumpulan *decission tree* yang pembangunan pohon berikutnya akan bergantung pada pohon sebelumnya. Pohon pertama dalam *XGboost* akan lemah dalam melakukan klasifikasi dengan inialisasi *probability* yang ditentukan oleh peneliti dan kemudian akan dilakukan update bobot pada setiap pohon yang dibangun sehingga menghasilkan kumpulan pohon klasifikasi yang kuat. Prediksi dilakukan dengan menjumlahkan seluruh bobot yang ada di setiap pohon dan kemudian memasukkan nilai tersebut ke fungsi logistic [8].

3. HASIL DAN PEMBAHASAN

a. Testing

Penelitian ini merupakan penelitian eksperimen. Pengumpulan data dalam penelitian ini meliputi studi literatur berupa buku, jurnal, dan karya ilmiah yang relevan dengan prediksi Peningkatan Jumlah Nasabah Deposito Berjangka dan dataset yang digunakan adalah Bank Marketing Dataset dari UCI Repository Machine Learning (<https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>), dimana terdapat 45211 record, 16 atribut dan 2 target kelas yang kemudian akan diolah menggunakan tool Google Colabs dan 80% data akan dibagi untuk data

training dan 20% untuk data testing dengan Python Programming menggunakan algoritma *K-Nearest Neighbor (KNN)*, *Decision Tree*, *Random Forest* dan *XGBoost*. Pada dataset terdapat 16 atribut antara lain : age, marital, default, housing, loan, month, duration, campaign, pdays, previous, emp.var.rate, cons.price.idx, cons.conf.idx, euribor3m, nr.employed dan y. Dalam uji coba ini, langkah-langkah untuk mendapatkan nilai akurasi adalah eksplorasi dataset, preprocessing, pemodelan dan evaluasi.



Gambar 1. Tahapan Penelitian

b. Hasil dan Pembahasan

Langkah-langkah yang dilakukan dalam penelitian ini adalah sebagai berikut:

a. *Data exploration* (Pengumpulan Data)

Dataset yang digunakan adalah data yang diperoleh dari UCI Repository. Dimana datanya terdiri dari 45211 record dengan 17 variabel atau atribut predictor.

b. *Data Pre-processing*

preprocessing adalah proses yang mengubah data mentah ke dalam bentuk yang lebih mudah dipahami. Proses ini penting dilakukan karena data mentah sering kali tidak memiliki format yang teratur. Selain itu, *data mining* juga tidak dapat memproses data mentah, sehingga data yang telah melalui proses *preprocessing* akan lebih mudah untuk diolah dalam pengujian

c. *Evaluation/Testing* (Pengujian)

Pengujian dilakukan dengan menggunakan algoritma *K-Nearest Neighbors (KNN)*, *Decision Tree*, *Random Forest* dan *XGBoost*. Masing-masing algoritma akan dilakukan pengujian menggunakan 80% *data training* dan 20% sebagai *data testing*.

d. Result (Hasil)

Berdasarkan penelitian yang telah dilakukan untuk mengetahui nilai akurasi dalam memprediksi Peningkatan Jumlah Nasabah Deposito Berjangka Menggunakan Algoritma *K-Nearest Neighbors* (KNN), *Decision Tree*, *Random Forest* Dan *XGBoost* menggunakan Bank Marketing Dataset dari UCI Repository Machine Learning diperoleh hasil sebagai berikut:

Tabel 1. Tabel Perbandingan Hasil Akurasi

Algoritma	Hasil Akurasi
<i>K-Nearest Neighbors</i> (KNN)	90,96%
<i>Decision Tree</i>	92,02%
<i>Random Forest</i>	90,68%
<i>XGBoost</i>	92,36%

Berdasarkan tabel perbandingan hasil akurasi diatas, nilai akurasi untuk model algoritma *K-Nearest Neighbors* (KNN) sebesar 90,96%, nilai akurasi algoritma *Decision Tree* sebesar 92,02%, nilai akurasi algoritma *Random Forest* sebesar 90,68% dan nilai akurasi algoritma *XGBoost* sebesar 92,36%.

4. KESIMPULAN

Dari hasil penelitian yang telah dilakukan algoritma *XGBoost* memiliki nilai akurasi yang lebih tinggi dibandingkan dengan menggunakan algoritma *K-Nearest Neighbors* (KNN), *Decision Tree* dan *Random Forest* yaitu sebesar 92,36%. Nilai akurasi untuk model algoritma *K-Nearest Neighbors* (KNN) sebesar 90,96%, nilai akurasi algoritma *Decision Tree* sebesar 92,02% dan nilai akurasi algoritma *Random Forest* sebesar 90,68%. Sehingga dapat disimpulkan bahwa penerapan algoritma *XGBoost* mampu menghasilkan tingkat akurasi dalam memprediksi Peningkatan Jumlah Nasabah Deposito Berjangka yang lebih baik dibandingkan menggunakan algoritma yaitu algoritma *K-Nearest Neighbors* (KNN), *Decision Tree* dan *Random Forest*.

REFERENSI

- [1] Apriyani, "Analisis Pengaruh Perubahan Inflasi, Tingkat Suku Bunga Sbi, Dan Ihsq Terhadap Dana Pihak Ketiga Masyarakat Pada Perbankan Indonesia", *Jurnal Manajemen Bisnis*, Vol. 10, No.2, pp. 193-204, 2021.
- [2] Yahya dan Hidayanti, "Penerapan Algoritma K-Nearest Neighbor Untuk KlasifikasiEfektivitas Penjualan Vape (Rokok Elektrik) pada Lombok Vape On", *Infotek : Informatika dan Teknologi*, Vol. 3, No.2, pp. 104-114, 2020.
- [3] Isman, Ahmad dan Latief, "Perbandingan Metode KNN Dan LBPH Pada Klasifikasi Daun Herbal", *Jurnal RESTI*, Vol. 5, No.3, pp. 557-564, 2021.
- [4] Sandag, G. A, "Model Prediksi Kemenangan Tim dalam Game League of Legend Menggunakan Algoritma Decision Tree", *Jurnal Komputer Terapan*, Vol. 7, No.1, pp. 42-52, 2021.
- [5] Rianto dan Yunis, "Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Random Forest", *Paradigma*, Vol. 23, No.1, 2021.
- [6] Irawan, Perkasa, Yurindra, Wahyuningsih dan Helmud, "Perbandingan Klassifikasi SMS Berbasis Support Vector Machine, Naive Bayes Classifier, Random Forest dan Bagging Classifier", *Jurnal SISFOKOM*, Vol. 10, No.3, pp. 432-437, 2021.
- [7] R. Chairunisa, Adiwijaya dan W. Astuti, "Perbandingan CART dan Random Forest untuk Deteksi Kanker berbasis Klasifikasi Data Microarray," *Jurnal Resti*, Vol. 4 No. 5, 2020.

- [8] M. Syukron, R. Santoso Dan T. Widiarih, "Perbandingan Metode Smote *Random Forest* Dan Smote Xgboost Untuk Klasifikasi Tingkat Penyakit Hepatitis C Pada *Imbalance Class Data*," Jurnal Gaussian, 2020.