

TUGAS AKHIR PENAMBANGAN DATA

**PREDIKSI PEMBELI E-READER DENGAN MENGGUNAKAN TEKNIK DATA MINING (CRISP-DM) BERDASARKAN DATASET PEMBELI**

DIUSULKAN OLEH

Andhieka Agrestya Al Ara Ab 1402020117

Muhammad Farhan Syach 1402020097

Ridho Gymnastiar Al-Rasyid 1402020069

Rio Griya Putra 1402020092

FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS YARSI

2022

**PENDAHULUAN**

1. **LATAR BELAKANG**

E-Reader adalah peralatan elektronik yang digunakan untuk membaca e-book. Produksi e-Reader perlu diprediksi dengan akurat, karena hasil prediksi yang akurat sangat penting untuk membuat kebijakan perdangan produk E-Reader.

Metode prediksi yang kelompok kami gunakan adalah Decision Tree yang merupakan salah satu cara data processing dalam memprediksi masa depan dengan cara membangun klasifikasi atau regresi model dalam bentuk struktur pohon. Kelompok kami juga menggunakan metode algoritma Random Forest dalam meningkatan nilai akurasi yang digunakan setelah algoritma Decision Tree.

Kelebihan algoritma yang kami gunakan itu sendiri untuk Decision Tree memiliki kelebihan dapat menemukan kombinasi data yang tidak terduga, memiliki akurasi yang baik, Intergrasi yang mudah kesistem basis data dan dapat menghilangkan perhitungan yang tidak diperlukan. Lalu untuk kekurangan Decision Tree yaitu kesulitan dalam merancang pohon keputusan yang optimal. Perhitungan jumlah kesalahan dari setiap level dalam pohon keputusan lumayan besar dan rawan terjadi tumpang tindih. Algoritma Random Forest yang kami gunakan disini juga membantu untuk menambahkan nilai akurasi akhirnya.

Penggunaan Decision Tree sangatlah rumit. Karena kita harus mempelajari cara memanggil dan menggambar tree nya sesuai kebutuhan data yang kita miliki. Tetapi dibalik sulitnya penggunaan algoritma Decision Tree, algoritma Decision Tree memiliki nilai akhir accuracy yang cukup tinggi.

1. **TUJUAN**
2. Membuat model prediksi produksi E-Reader menggunakan algoritma Decision Tree dan Random Forest
3. Melakukan evaluasi model prediksi E-Reader

**BAB II. METODE**

1. **Decision Tree Algorithm dan Random Forest Classifier**
   1. **Decision Tree Algorithm**

Decision Tree merupakan salah satu cara data processing dalam memprediksi masa depan dengan cara membangun klasifikasi atau regresi model dalam bentuk struktur pohon.

* 1. **Random Forest Cluster**

Random forest adalah suatu algoritma yang digunakan untuk klasifikasi data dalam jumlah yang besar. Random forest merupakan kombinasi dari masing – masing pohon (tree) dari model Decision Tree yang baik, dan kemudian dikombinasikan ke dalam satu model.

1. **Desain Eksperimen**

Desain eksperimen untuk prediksi produksi e-Reader menggunakan Decision Tree Algorithm



Gambar 1. Desain Eksperimen menggunakan Decision Tree

Tahapan awal seletah dataset didapatkan (ROOT Node) adalah decision node. Pada tahapan ini dataset dibagi menjadi dua yang sebelumnya permasalahan seperti missing values, outliers, inconsistent data sudah diatasi serta sudah dilakukan normalisasi (memfilter atribut yang tidak diperlukan). Data pengujian ini menggunakan … ‘gini’. Setelah itu melakukan prediksi pada data train lalu membandingkan akurasi antara 2 data tersebut.

Desain Eksperimen untuk prediksi produksi e-Reader menggunakan Random Forest Classifier.



Gambar 2. Desain eksperimen Random Forest Classifier

1. **Dataset**

Dataset yang digunakan pada proyek akhir ini adalah …., lampirkan penjelasan mengenai jumlah sampel, atribut, nilai yang diprediksi / estimasi. lampirkan link untuk mengakses dataset tersebut.

1. **Metode Evaluasi**

Metode evaluasi yang digunakan adalah nilai mean squared error. Adapun rumus mean squared error seperti di bawah ini

Schematic

Description automatically generated

**BAB III. PEMBAHASAN**

1. **Hasil dan Analisis**

Berdasarkan hasil eksperimen yang dilakukan, yaitu mengimplementasian algoritma Decision Tree dan Random Forest pada metode Classification e-Reader Adoption berdasarkan perbandingan antara dua data yaitu data training dan data testing.

A black screen with white text

Description automatically generated with low confidence

Gambar 3.1. Atribut Dataset.

Gambar 3 merupakan atribut awal dari dataset yang kami pakai. Setelah itu, kami membagi dataset menjadi 2 yaitu, data train dan data test. Lalu, kami menghilangkan beberapa atribut yang tidak digunakan atau tidak diperlukan dalam memprediksi penjualan e-reader.

Text

Description automatically generated

Gambar 3.2. Atribut data setelah dilakukan drop atribut.

Setelah menghilangkan beberapa atribut, selanjutnya membuat model Decision Tree Classifier dengan menggunakan metode “gini”



Diagram

Description automatically generatedGambar 3.3. Skor akurasi dengan “gini”

Gambar 3.4. Tree Plot

Diagram

Description automatically generated

Gambar 3.5. Tree plot with Graphviz.

1. **Kesimpulan**

Implementasi Decision Tree Classifier dalam memprediksi pembelian e-Reader dapat dilakukan. Berdasarkan hasil eksperimen, kami dapat memprediksi kapan pembeli akan membeli e-Reader. Kita dapat memprediksi dengan menggunakan algoritma prediksi lain seperti Random Forest dan SVM.

1. **Saran dan Rencana Kedepan**

Untuk menambah performa dari model prediksi e-Reader, kedepannya kami akan menambah jumlah data pelatihan, melakukan eksplorasi penerapan algoritma optimasi lain seperti Support Vector Machine.

**DAFTAR PUSTAKA**

1. https://sis.binus.ac.id/2022/01/21/decision-tree-algoritma-beserta-contohnya-pada-data-mining/ – Pengertian Decision Tree
2. http://learningbox.coffeecup.com/05\_2\_randomforest.html – Pengertian Random Forest