

**LAPORAN KLASIFIKASI PADA AUCTION VERIFICATION
MENGUNAKAN NAÏVE BAYES**

**Disusun guna memenuhi tugas mata kuliah
Pengantar Kecerdasan Buatan**



Oleh:

Bintang Rizky Zulfikar	1301223104
Muhammad Naufal Arifin	1301220356
Sakie Giano Suza	1301223205

**KELAS IF-46-12
PROGRAM STUDI INFORMATIKA
FAKULTAS INFORMATIKA
TELKOM UNIVERSITY
TAHUN 2024**

PENDAHULUAN

Auction Verification adalah dataset yang digunakan untuk studi ilmiah, tujuan dataset ini sendiri adalah mengurangi biaya dalam verifikasi hasil simultaneous multi-round auctions (SMRA). Setiap instance dari dataset mewakili satu proses verifikasi. Verifikasi ialah apakah harga tertentu mungkin untuk produk tertentu dan apakah penawar tertentu memenangkan produk dengan harga tersebut. Penjelasan dari variabel yang ada pada Dataset:

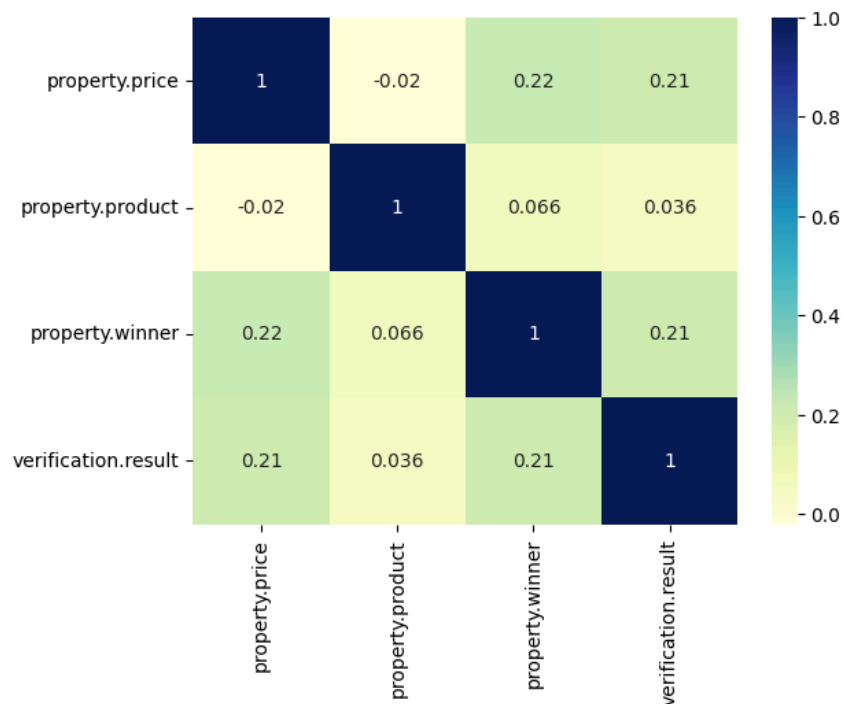
1. process.b1.capacity = Kapasitas (jumlah maksimal produk yang menang) dari bider 1. (variabel ini tidak akan dipakai dalam model).
2. process.b2.capacity = Kapasitas (jumlah maksimal produk yang menang) dari bider 2. (variabel ini tidak akan dipakai dalam model).
3. process.b3.capacity = Kapasitas (jumlah maksimal produk yang menang) dari bider 3. (variabel ini tidak akan dipakai dalam model).
4. process.b4.capacity = Kapasitas (jumlah maksimal produk yang menang) dari bider 4. (variabel ini tidak akan dipakai dalam model).
5. property.price = Harga saat ini diverifikasi.
6. property.product = Produk saat ini diverifikasi.
7. property.winner = Penawar saat ini diverifikasi sebagai pemenang produk (0 jika hanya harga yang diverifikasi).
8. verification.result = Hasil verifikasi biner (apakah hasil yang diverifikasi mungkin?).
9. verification.time = Runtime prosedur verifikasi. (variabel ini tidak akan dipakai dalam model).

Target variabel yang akan digunakan adalah verification.result, penggunaan variabel ini dikarenakan tujuan dari setiap instance adalah memeriksa apakah hasil yang diverifikasi oleh property.price, property.product dan property.winner mungkin. Variabel

verification.result adalah False atau True (biner), oleh sebab itu lebih tepat apabila mengambil studi kasus klasifikasi.

Alasan variabel verification.time, process.b1.capacity, process.b2.capacity, process.b3.capacity, process.b4.capacity tidak digunakan bahkan dihapus dalam model adalah karena atribut tersebut tidak terlalu penting dalam pengambilan kesimpulan dari setiap instance dan untuk runtime sangat bergantung dengan device yang digunakan sehingga runtime akan sangat bervariasi sesuai dengan spesifikasi komputer.

Pengolahan dataset kami menggunakan library pandas. Dataset yang awalnya memiliki variabel verification.time, process.b1.capacity, process.b2.capacity, process.b3.capacity, process.b4.capacity akan dihapus karena alasan sebelumnya. Dataset tidak memiliki missing value, duplicate data. Korelasi dari masing-masing variabel dapat dilihat dari heatmap ini:



Sebagian besar variabel memiliki korelasi rendah atau hampir tidak ada korelasi satu sama lain, menunjukkan bahwa mereka mungkin beroperasi secara relatif independen dalam konteks dataset ini.

Dataset dilakukan splitting dengan pengambilan secara random data latih sebanyak 80% dari data yang ada, sedangkan untuk data uji adalah data sisa dari hasil pengambilan data oleh data latih.

METODE

Metode yang akan digunakan adalah Naïve Bayes dikarenakan variabel target yang digunakan adalah sebuah binary classification (False/True). Kenapa metode ini lebih cocok dari pada metode lain? Naïve Bayes merupakan metode pembuatan model classification yang dalam pembuatan model nya tidak mepedulikan correlation dari satu atribut dengan atribut lainnya atau bisa dibilang menganggap semua atribut adalah atribut yang independen, ini sesuai dengan kesimpulan dari pembacaan heatmap yang telah dipaparkan dimana setiap variabel memiliki correlation yang lemah dengan variabel lainnya, membuktikan bahwa setiap variabel dari data yang dipakai adalah independen. Jenis Naïve Bayes yang dipakai adalah for Naïve Bayes for Continuous-valued Features karena atribut dari dataset adalah integer bukan label.

Parameter yang akan dipakai terdapat Mean (rata-rata), Standard Deviation, Gaussian Formula (digunakan saat testing), hasil model baik itu False atau True, data latih, dan data uji. Mean dan standard deviation itu digunakan dalam pembuatan model dari naive bayes. Gaussian Formula adalah rumus yang dipakai sebagai bahan pengambil keputusan berdasarkan model yang telah dimiliki.

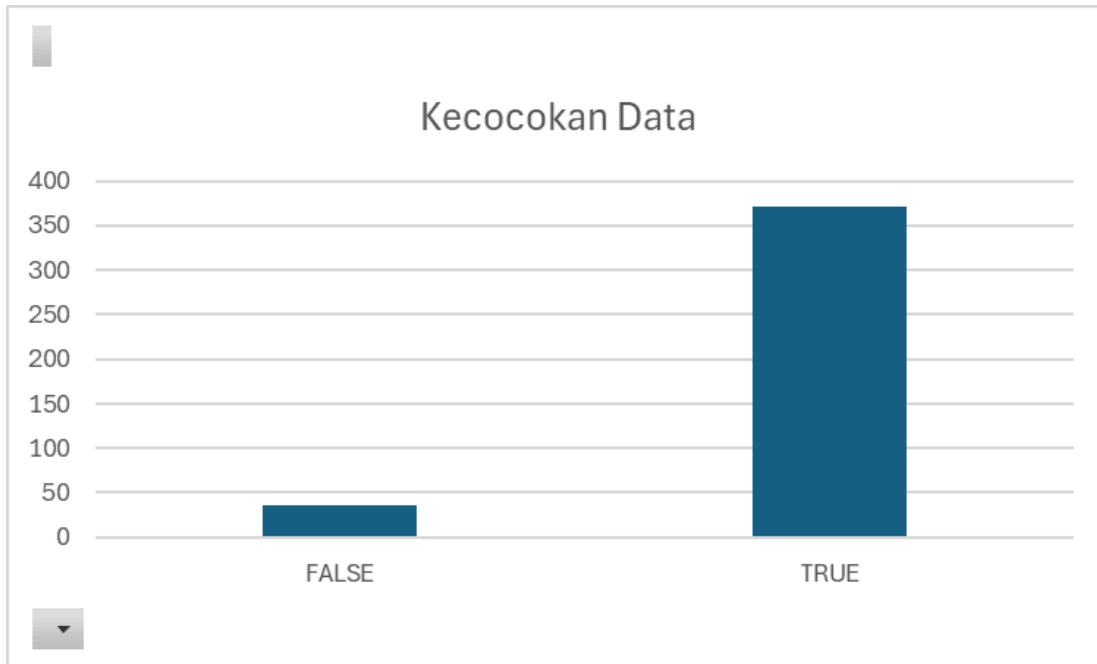
HASIL DAN ANALISIS

Model yang didapat setelah menjalankan proses learning data:

	property.price	property.product	property.winner	verification.result
mean	76,0444444444	3,4577777778	1,08	TRUE
standar deviasi	9,05823312	1,731856055	1,22212	

	property.price	property.product	property.winner	verification.result
mean	70,8161816891	3,2114975160	0,366217175	FALSE
standar deviasi	7,76086990	1,812396890	1,01936	

Graph yang menyatakan kecocokan dengan data asli:



Dilihat dari graph diatas dapat dilihat true adalah data yang cocok antara data test real dengan data hasil testing, dan false adalah data yang tidak cocok antara data test real dengan data hasil evaluation. Setelah model Naïve Bayes untuk klasifikasi biner dijalankan pada dataset yang telah diproses, didapatkan hasil kecocokan data hasil testing dengan data yang asli sebanyak 90.95%. Artinya, model mampu memprediksi hasil verifikasi dengan benar sebanyak 90.95% dari total data uji yang digunakan.

Akurasi sebesar 90.95% menunjukkan bahwa model Naïve Bayes yang digunakan memiliki performa yang sangat baik dalam memprediksi hasil verifikasi (True/False) berdasarkan variabel input yang ada. Ini berarti model dapat mengidentifikasi hasil yang benar di sebagian besar kasus.

KESIMPULAN

Model Naïve Bayes yang digunakan menunjukkan performa yang sangat baik dengan akurasi 90.95% dalam memprediksi hasil verifikasi simultaneous multi-round auctions (SMRA). Pemilihan fitur dan metode yang tepat berdasarkan analisis korelasi

serta asumsi independensi variabel telah terbukti efektif. Penghapusan variabel yang tidak signifikan dan memastikan dataset dalam kondisi bersih berkontribusi pada keberhasilan model ini.