

ABSTRACT

Melvin Lee (01671220001)

COMPARISON OF BAGGING, BOOSTING, AND STACKING ENSEMBLE MODELS FOR AIRLINE CUSTOMER SATISFACTION ANALYSIS
(xiii + 59 pages; 27 figures; 0 tables, 0 diagrams, 0 Appendix)

With the end of COVID-19 pandemic and subsequent lockdowns last year, air travel has soared high, with an increase of 30.1% compared to last year according to one report. The rise of number of passengers means a good opportunity for the airline carriers to recoup losses due to lockdowns, and competition becomes heated as rival carriers try to lure new and old customers into their services. To remain competitive, more and more companies are turning towards machine learning to analyze large amounts of data to gain an edge towards their competitors, with ensemble learning being one of the many methods employed for the analysis work.

In this study, Decision Tree, Random Forest, Boosting, and Stacking methods will be chosen for comparative study, which will be supplied with Airline Satisfaction dataset which is cleaned of null values and changing data types, for the study itself and then compared with each other using confusion matrix, precision-recall-f1-score accuracy metrics, ROC curve, and feature importances.

The results have shown that while the three chosen classifiers are almost similar in their overall success rate, with Bagging method reaching 96.117%, Boosting with a rate of 96.037%, and stacking with a rate of 96.264%, overall Stacking has the highest rate among all. These results show the almost negligible differences on all three main ensemble learning methods in terms of efficacy. Additional studies with larger datasets, and more varieties of ensemble learning methods can improve the overall judgement of the results.

Keywords: ensemble learning, airline satisfaction, bagging, boosting, stacking

References 43 (2011 – 2023)

ABSTRAK

Melvin Lee (01671220001)

PERBANDINGAN BAGGING, BOOSTING, DAN STACKING MODEL ENSEMBLE LEARNING UNTUK ANALISIS KEPUASAN PELANGGAN MASKAPAI

(xiii + 59 halaman; 247 gambar; 0 tabel; 0 diagram; 0 lampiran)

Dengan akhiran pandemi *COVID-19* dan *lockdown* yang terjadi tahun lalu, perjalanan udara melonjak tinggi, dengan peningkatan sebesar 30,1% dibandingkan tahun lalu menurut sebuah laporan. Peningkatan jumlah penumpang berarti peluang bagus bagi maskapai penerbangan untuk menutup kerugian akibat *lockdown*, dan persaingan menjadi memanas ketika maskapai pesaing mencoba memikat pelanggan baru dan lama untuk menggunakan layanan mereka. Agar tetap kompetitif, semakin banyak perusahaan yang beralih ke pembelajaran mesin untuk menganalisis data dalam jumlah besar guna mendapatkan keunggulan dibandingkan pesaing mereka, dengan pembelajaran ansambel menjadi salah satu dari banyak metode yang digunakan untuk pekerjaan analisis.

Dalam studi ini, metode Decision Tree, Random Forest, Boosting, dan Stacking akan dipilih untuk studi komparatif, yang akan dilengkapi dengan dataset Kepuasan Maskapai yang dibersihkan dari nilai null dan tipe data yang berubah, untuk studi itu sendiri dan kemudian dibandingkan dengan masing-masing metode. lainnya menggunakan matriks konfusi, metrik akurasi skor recall-f1, kurva ROC, dan kepentingan fitur.

Hasilnya menunjukkan bahwa meskipun ketiga pengklasifikasi yang dipilih memiliki tingkat keberhasilan keseluruhan yang hampir serupa, dengan metode Bagging mencapai 96,117%, Boosting dengan tingkat 96,037%, dan penumpukan dengan tingkat 96,264%, secara keseluruhan Penumpukan memiliki tingkat tertinggi di antara pengklasifikasi lainnya. semua. Hasil ini menunjukkan perbedaan yang hampir dapat diabaikan pada ketiga metode pembelajaran ansambel utama dalam hal kemanjuran. Studi tambahan dengan kumpulan data yang lebih besar, dan lebih banyak variasi metode pembelajaran ansambel dapat meningkatkan penilaian hasil secara keseluruhan.

Kata Kunci: *ensemble learning, airline satisfaction, bagging, boosting, stacking*

Referensi: 43 (2011 – 2023)