



**A COMPARATIVE STUDY: APPLICATION OF PRINCIPAL
COMPONENT ANALYSIS AND RECURSIVE FEATURE
ELIMINATION IN MACHINE LEARNING FOR STROKE
PREDICTION**

Skripsi

**Untuk memenuhi persyaratan
Dalam menyelesaikan Sarjana Strata-1 Ilmu Komputer**

Oleh

ARYA SYIFA HERMIATI

NIM 2011016220035

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

JUNI 2024



**A COMPARATIVE STUDY: APPLICATION OF PRINCIPAL
COMPONENT ANALYSIS AND RECURSIVE FEATURE
ELIMINATION IN MACHINE LEARNING FOR STROKE
PREDICTION**

Skripsi

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

ARYA SYIFA HERMIATI

NIM 2011016220035

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2024**

SKRIPSI

A COMPARATIVE STUDY: APPLICATION OF PRINCIPAL COMPONENT ANALYSIS AND RECURSIVE FEATURE ELIMINATION IN MACHINE LEARNING FOR STROKE PREDICTION

Oleh :

ARYA SYIFA HERMIATI
NIM. 2011016220035

Telah dipertahankan di depan Dosen Penguji pada tanggal 24 Juni 2024.

Susunan Dosen Penguji:
Pembimbing Utama



Rudy Herteno, S.Kom., M.Kom.

NIP. 198809252022031003

Penguji I



Triando Hamonangan Saragih, S.Kom, M.Kom

NIP. 199308242019031012

Pembimbing Pendamping



Fatma Indriani, S.T., M.I.T., Ph.D.

NIP. 198404202008122004

Penguji II



Muliadi, S.Kom, M.Sc

NIP. 197804222010121002



Banjarnegara, 26 Juni 2024
Koordinator PS Ilmu Komputer

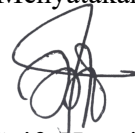
Wan Budiman, S.T., M.Kom.

NIP. 197703252008121001

SURAT PERNYATAAN

Dengan ini saya menyatakan bahwa dalam jurnal ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 26 Juni 2024
Yang Menyatakan,



Arya Syifa Hermiati

NIM. 2011016220035

ABSTRAK

SEBUAH STUDI PERBANDINGAN: PENERAPAN *PRINCIPAL COMPONENT ANALYSIS* DAN *RECURSIVE FEATURE ELIMINATION* DALAM *MACHINE LEARNING* UNTUK PREDIKSI STROKE

(Oleh: Arya Syifa Hermiati; Pembimbing: Rudy Herteno, S.Kom., M.Kom., Fatma Indriani, S.T., M.I.T., Ph.D.;2024;halaman)

Stroke adalah penyakit yang terjadi pada otak dan dapat menyebabkan disfungsi otak baik vokal maupun global. Penelitian stroke terutama bertujuan untuk memprediksi risiko dan kematian. *Machine learning* dapat digunakan untuk mendiagnosa dan memprediksi penyakit di bidang kesehatan, terutama dalam prediksi stroke. Namun, mengumpulkan data rekam medis untuk memprediksi suatu penyakit biasanya menimbulkan banyak *noise* karena tidak semua variabel penting dan relevan dengan proses prediksi. Dalam hal ini, reduksi dimensi sangat penting untuk menghilangkan fitur-fitur yang berisik (yaitu, tidak relevan) dan berlebihan. Penelitian ini bertujuan untuk memprediksi stroke dengan menggunakan *Recursive Feature Elimination* sebagai seleksi fitur, *Principal Component Analysis* sebagai ekstraksi fitur, dan kombinasi dari *Recursive Feature Elimination* dan *Principal Component Analysis*. Dataset yang digunakan dalam penelitian ini adalah prediksi stroke dari Kaggle. Metodologi penelitian terdiri dari *pre-processing*, SMOTE, *10-fold Cross-Validation*, seleksi fitur, ekstraksi fitur, dan *machine learning*, yang meliputi SVM, *Random Forest*, *Naive Bayes*, dan *Linear Discriminant Analysis*. Dari hasil yang didapatkan, SVM dan *Random Forest* mendapatkan nilai akurasi tertinggi yaitu 0.8775 dan 0.9511 tanpa menggunakan PCA dan RFE, *Naive Bayes* mendapatkan nilai tertinggi yaitu 0.7685 ketika melalui PCA dengan seleksi 20 fitur dilanjutkan seleksi fitur RFE dengan seleksi 5 fitur, dan LDA mendapatkan akurasi tertinggi dengan 20 fitur dari seleksi fitur dan dilanjutkan ekstraksi fitur dengan nilai 0.7963. Dapat disimpulkan dalam penelitian ini bahwa SVM dan *Random Forest* mendapatkan nilai akurasi tertinggi tanpa teknik PCA dan RFE, sedangkan *Naive Bayes* dan LDA menunjukkan performa yang lebih baik dengan menggunakan kombinasi teknik PCA dan RFE. Implikasi dari penelitian ini adalah untuk mengetahui pengaruh RFE dan PCA pada *machine learning* untuk meningkatkan prediksi stroke.

Kata kunci: *Recursive Feature Elimination, Principal Component Analysis, Support Vector Machine, Random Forest, Naive Bayes, Linear Discriminant Analysis*

ABSTRAK

A COMPARATIVE STUDY: APPLICATION OF PRINCIPAL COMPONENT ANALYSIS AND RECURSIVE FEATURE ELIMINATION IN MACHINE LEARNING FOR STROKE PREDICTION

(By: Arya Syifa Hermiati; Supervisors: Rudy Herteno, S.Kom., M.Kom., Fatma Indriani, S.T., M.I.T., Ph.D.;2024;page)

Stroke is a disease that occurs in the brain and can cause both vocal and global brain dysfunction. Stroke research mainly aims to predict risk and mortality. Machine learning can be used to diagnose and predict diseases in the healthcare field, especially in stroke prediction. However, collecting medical record data to predict a disease usually makes much noise because not all variables are important and relevant to the prediction process. In this case, dimensionality reduction is essential to remove noisy (i.e., irrelevant) and redundant features. This study aims to predict stroke using Recursive Feature Elimination as feature selection, Principal Component Analysis as feature extraction, and a combination of Recursive Feature Elimination and Principal Component Analysis. The dataset used in this research is stroke prediction from Kaggle. The research methodology consists of pre-processing, SMOTE, 10-fold Cross-Validation, feature selection, feature extraction, and machine learning, which includes SVM, Random Forest, Naive Bayes, and Linear Discriminant Analysis. From the results obtained, the SVM and Random Forest get the highest accuracy value of 0.8775 and 0.9511 without using PCA and RFE, Naive Bayes gets the highest value of 0.7685 when going through PCA with selection of 20 features followed by RFE feature selection with selection of 5 features, and LDA gets the highest accuracy with 20 features from feature selection and continued feature extraction with a value of 0.7963. It can be concluded in this study that SVM and Random Forest get the highest accuracy value without PCA and RFE techniques, while Naive Bayes and LDA show better performance using a combination of PCA and RFE techniques. The implication of this research is to know the effect of RFE and PCA on machine learning to improve stroke prediction.

Keywords: Recursive Feature Elimination, Principal Component Analysis, Support Vector Machine, Random Forest, Naive Bayes, Linear Discriminant Analysis

KATA PENGANTAR

Puji syukur kepada ALLAH SWT yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “*A Comparative Study: Application of Principal Component Analysis and Recursive Feature Elimination in Machine Learning for Stroke Prediction*” dengan baik untuk memenuhi syarat dalam menyelesaikan pendidikan program Strata-1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Oleh sebab itu, penulis ucapkan terima kasih kepada:

1. Allah SWT, segala puji syukur atas segala kemudahan yang diberikan oleh-Nya sehingga akhirnya laporan magang ini dapat diselesaikan dengan baik dan tanpa hambatan yang berarti.
2. Mama, Abah, Oca, Nini, Acil Huda dan seluruh keluarga penulis yang penulis cintai dan telah memberi dukungan material serta moral kepada penulis sehingga penulis mendapatkan dorongan untuk lebih bersemangat menyelesaikan skripsi setiap harinya.
3. Bapak Rudy Herteno, S.Kom., M.Kom., selaku Dosen Pembimbing Utama yang selalu memberikan masukan dan bantuan dalam menyelesaikan penelitian skripsi penulis.
4. Ibu Fatma Indriani, S.T., M.I.T., Ph.D., selaku Dosen Pembimbing Pendamping yang selalu memberikan penjelasan dan bantuan dalam menyelesaikan penulisan naskah jurnal dan skripsi.
5. Bapak Triando Hamonangan Saragih, S.Kom., M.Kom dan Muliadi, S.Kom., M.Cs. selaku dosen ketua penguji dan anggota penguji skripsi.
6. Bapak Irwan Budiman, S.T., M.Kom. selaku ketua program studi Ilmu Komputer beserta seluruh dosen dan karyawan/staff pegawai Fakultas MIPA, Universitas Lambung Mangkurat atas bantuan yang diberikan selama penulis mengikuti studi.
7. Rizqi Nur Abidah, sebagai sahabat penulis yang selalu siap mendengarkan

keluh kesah penulis selama ini, teman bertukar pikiran, memberikan masukan serta tempat penulis bercakap random dikala kerumitan skripsi melanda.

8. Yunita Sari, Helma Ghinaya, Nor Aida, Mita Azzahra Ramadhan, Adinda Ayu Puspita, Vania Zaliany, Noor Zalekha, Putri Nabela, dan semua ciway yang sudah membantu penulis berproses selama pengerjaan skripsi dan jurnal.
9. Keluarga Pentol yang selalu menyemangati dan mendoakan penulis untuk kelancaran pengerjaan skripsi.
10. Teman teman Ilmu Komputer angkatan 2020 terima kasih untuk canda, tawa, perjuangan yang sudah dilewati bersama, untuk semua kenangan manis yang telah terukir selama ini. Senang bisa menjadi salah satu bagian dari kehidupan kalian.
11. Wave to Earth yang lagunya selalu penulis dengarkan selama pengerjaan skripsi dan jurnal.
12. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, maka saran dan kritik dari semua pihak sangat diharapkan demi penyempurnaan pada penelitian selanjutnya. Diharapkan skripsi ini dapat bermanfaat bagi semua pihak dan dapat menambah pengetahuan kita semua.

Banjarbaru, 26 Juni 2024



Penulis