

INTISARI

PENDEKATAN *EXPLAINABLE MACHINE LEARNING* DALAM PEMODELAN DATA SURVIVAL SEBAGAI ALTERNATIF MODEL *COX* *PROPORTIONAL HAZARDS* (STUDI KASUS: DATA *VETERAN LUNG* *CANCER*)

Oleh

Syifa Azzahra Susilo

22/496431/PA/21342

Cox Proportional Hazards (CPH) merupakan metode standar dalam analisis survival untuk mengestimasi risiko kematian pasien. Namun, akurasi model ini sering kali terbatas ketika data melanggar asumsi *proportional hazards* atau memiliki hubungan kovariat yang bersifat non-linear. Penelitian ini bertujuan untuk mengevaluasi potensi pendekatan *machine learning* sebagai alternatif yang lebih efektif dalam mengatasi keterbatasan tersebut, dengan studi kasus pada dataset *Veteran Lung Cancer*. Empat algoritma dibandingkan kinerjanya, yaitu *Cox Proportional Hazards*, *Random Survival Forest* (RSF), *Survival Support Vector Machine* (SSVM), dan *XGBoost Survival*.

Kinerja model dievaluasi menggunakan *Concordance Index* (*C-index*), sedangkan aspek interpretabilitas model *black box* dijelaskan menggunakan metode *SHapley Additive exPlanations* (SHAP). Hasil penelitian menunjukkan bahwa data *Veteran Lung Cancer* secara signifikan melanggar asumsi proporsionalitas ($p < 0.005$). Dalam kondisi tersebut, model RSF mampu mengungguli CPH dengan nilai *C-index* sebesar 0.7812, lebih tinggi dibandingkan dengan CPH sebesar 0.7618. Interpretasi berbasis SHAP terbukti konsisten dengan temuan statistik pada model *Cox*, serta memberikan nilai tambah dengan mengungkap pola non-linear dan interaksi kompleks antarvariabel. Penelitian ini menyimpulkan bahwa integrasi RSF dan SHAP merupakan alternatif yang valid dan lebih akurat untuk menganalisis dataset *Veteran Lung Cancer* tanpa mengorbankan aspek interpretabilitas model.

ABSTRACT

EXPLAINABLE MACHINE LEARNING APPROACH IN SURVIVAL DATA MODELING AS AN ALTERNATIVE TO THE COX PROPORTIONAL HAZARDS MODEL (CASE STUDY: VETERAN LUNG CANCER DATA)

By

Syifa Azzahra Susilo

22/496431/PA/21342

Cox Proportional Hazards (CPH) is a standard method in survival analysis for estimating patient mortality risk. However, the accuracy of this model is often limited when the data violate the proportional hazards assumption or exhibit non-linear relationships among covariates. This study aims to evaluate the potential of machine learning approaches as more effective alternatives to address these limitations, using the Veteran Lung Cancer dataset as a case study. Four algorithms were compared, namely Cox Proportional Hazards, Random Survival Forest (RSF), Survival Support Vector Machine (SSVM), and XGBoost Survival.

Model performance was evaluated using the Concordance Index (C-index), while the interpretability of black-box models was addressed using the SHapley Additive exPlanations (SHAP) method. The results indicate that the Veteran Lung Cancer data significantly violate the proportional hazards assumption ($p < 0.005$). Under these conditions, the RSF model outperformed the CPH model, achieving a C-index of 0.7812 compared to 0.7618 for CPH. SHAP-based interpretations were consistent with the statistical findings of the Cox model and provided additional insights by revealing non-linear patterns and complex interactions among variables. This study concludes that the integration of RSF and SHAP constitutes a valid and more accurate alternative for analyzing the Veteran Lung Cancer dataset without compromising model interpretability.