

## INTISARI

### **METODE DATA SPLITTING STACKING DENGAN SUPPORT VECTOR MACHINE (SVM) DAN K-NEAREST NEIGHBORS (KNN) UNTUK PREDIKSI BANJIR DI WILAYAH KHUSUS IBUKOTA JAKARTA**

Oleh

M Haidar Azhar  
18/429291/PA/18682

Banjir selalu menjadi jenis bencana alam yang paling sering terjadi di Indonesia. Oleh karena itu, pemerintah dan para ahli berupaya untuk mengurangi kerugian akibat peristiwa tersebut dengan memperkirakan apakah akan terjadi banjir atau tidak. Berbagai metode dapat digunakan untuk memprediksi manifestasi banjir, salah satunya adalah dengan menggunakan *ensemble* yang disebut *stacked generalization* atau *stacking*. Meskipun studi tentang *stacking* cukup umum, diyakini bahwa metode *data splitting* masih sangat sedikit yang tidak diketahui untuk digunakan. Dengan demikian, teknik *data splitting stacking* dapat memperbaiki model prediksi menjadi lebih akurat.

*Stacking* dilakukan dengan menggunakan dua model sebagai dua pembelajar utama, *Support Vector Machine* (SVM) dan *K-Nearest Neighbors* (KNN). Kedua model ini dilatih dan diuji dengan menggunakan data data iklim Jakarta Indonesia dari tahun 2017 hingga 2021 dan data curah hujan tahunan tambahan dari provinsi Kerala di India, dengan kondisi iklim yaitu suhu minimum, suhu maksimum, suhu rata-rata, kelembaban rata-rata, curah hujan, kecepatan angin maksimum, dan kecepatan angin rata-rata sebagai prediktor. Selain itu, perbandingan tanpa *data splitting* juga dilakukan untuk membedakan kedua metode tersebut.

*Stacked generalization* menggunakan dataset Jakarta dan Kerala dengan pemisahan data menghasilkan akurasi masing-masing sebesar 95,62 dan 90,90. Sebaliknya, metode non-splitting menghasilkan akurasi sebesar 93,16 untuk dataset Jakarta dan 95,83 untuk dataset Kerala. Hasil menunjukkan bahwa *stacked generalization* menunjukkan akurasi yang lebih tinggi dan berkinerja lebih baik daripada SVM dan KNN untuk metode *splitting* dan *non-splitting* dengan metode *data splitting stacking* meningkatkan akurasi lebih signifikan daripada pendekatan *non-splitting*.

**Kata Kunci:** Prediksi banjir, Data iklim, *Data splitting*, *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Stacked Generalization*

## ABSTRACT

### DATA SPLITTING METHOD OF STACKING WITH SUPPORT VECTOR MACHINE (SVM) AND K-NEAREST NEIGHBORS (KNN) FOR PREDICTING FLOOD IN SPECIAL CAPITAL REGION OF JAKARTA

By

M Haidar Azhar  
18/429291/PA/18682

Flood occurrence has always been the most common type of natural disaster that frequently happens in Indonesia. Therefore, government and experts attempt to lower the damages of the event by predicting whether flood is happening or not. Various methods can be used to predict the manifestation of flood, one method is by using ensemble called stacked generalization or stacking. Although studies regarding stacking is quite common, it is believed that the data splitting method is still very much little to unknown to be used. Thus, the data splitting stacking technique can improve the prediction model to be more accurate.

The stacking is done by using two models as the two main learners, Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). These two models train and test by employing the data of Indonesia's Jakarta climate data from 2017 to 2021 and an additional yearly rainfall data from Kerala province in India, with climate conditions that are minimum temperature, maximum temperature, average temperature, average humidity, rainfall, maximum wind speed, and average wind speed as predictors. Moreover, comparison without data splitting is also conducted to differentiate both methods.

Stacked generalization using Jakarta and Kerala dataset with data splitting resulted an accuracy of 95.62 and 90.90 of accuracy, respectively. In contrast, the method of non-splitting produced an accuracy of 93.16 for Jakarta dataset and 95.83 for Kerala dataset. The result shows that stacked generalization has higher accuracy that indicated that it performs better than SVM and KNN for both splitting and non-splitting method with the splitting method stacking increased the accuracy more significantly rather than the non-splitting approach.

**Keywords:** Flood prediction, Climate data, Data splitting, Support Vector Machine, K-Nearest Neighbors, Stacked Generalization