

## ABSTRAK

Winly Williamdy (01032170001)

### **IMPLEMENTASI MACHINE LEARNING UNTUK ANALISIS KLASTER DAN ANALISIS PREDIKTIF DATA BPJS KESEHATAN INDONESIA TAHUN 2015-2016**

Skripsi, Fakultas Sains dan Teknologi (2021)

(xvi + 111 halaman; 42 gambar; 23 tabel; 5 lampiran)

Layanan kesehatan BPJS Kesehatan mengalami defisit sekitar 13 triliun Rupiah pada tahun 2019. Dengan tingginya angka defisit tersebut, pihak BPJS Kesehatan menaikkan jumlah iuran untuk seluruh kelas hingga sekitar 2 kali lipat iuran awal, namun hal tersebut mendapatkan penolakan dari berbagai pihak. Dengan memanfaatkan data yang telah disebar oleh pihak BPJS Kesehatan dan telah berkembangnya teknologi *machine learning*, maka dapat dilakukan analisis berupa analisis deskriptif dan analisis prediktif. Dengan menggunakan algoritma *K-Means Clustering*, peserta – peserta BPJS Kesehatan Indonesia dapat diklasterisasi sesuai dengan tingkah lakunya menurut analisis *RFM (Recency, Frequency, dan Monetary)*. Berdasarkan analisis menggunakan *elbow method*, didapatkan jumlah kluster optimum sebanyak 4 kluster. Keseluruhan data yang sekarang terpisah menjadi empat kluster kini terlihat memiliki karakteristik yang berbeda – beda. Penelitian ini juga mencoba untuk memberikan solusi terhadap permasalahan dengan melakukan prediksi jumlah klaim yang akan terjadi pada periode tertentu. Hasil prediksi yang dilakukan dengan beberapa model tidaklah akurat dan memiliki nilai *RMSE (Root Mean Square Error)* yang buruk yaitu berkisar pada nilai 0,0113 sampai 0,00325. Hal ini dikarenakan peristiwa orang berkunjung ke fasilitas kesehatan bersifat *random walk* dan tidak dapat diprediksi. Hasil dari analisis menggunakan model *ARIMA (AutoRegressive Integrated Moving Average)* menunjukkan bahwa jumlah klaim yang terjadi di setiap bulan tidaklah berkaitan satu sama lain karena tidak adanya nilai *PACF (Partial Autocorrelation Function)* dan *ACF (Autocorrelation Function)* yang signifikan. Oleh karena itu, model terbaik yang dapat memprediksi peristiwa ini adalah model *ARIMA* dengan orde (0,1,0) atau *random walk model with drift* di mana model ini memiliki nilai *RMSE* terkecil di antara model – model lain dalam penelitian ini, yaitu 0,003248. Kesimpulan terbaik yang dapat diberikan pada penelitian ini adalah menaikkan nilai iuran kepada beberapa peserta sesuai dengan kluster mereka. Kluster yang dinilai dapat menanggung beban ini adalah kluster 1 karena kebanyakan peserta dalam kluster ini adalah peserta yang berusia produktif dan berjenis kelamin pria.

Kata Kunci : BPJS Kesehatan Indonesia, *K-Means Clustering, Linear Regression, Random Forest Regressor, ARIMA, Bühlmann-Straub Credibility Theory*

Referensi : [21] (1987 – 2020)

## ABSTRACT

Winly Williamdy (01032170001)

### **IMPLEMENTATION OF MACHINE LEARNING FOR CLUSTER ANALYSIS AND PREDICTIVE ANALYSIS IN BPJS KESEHATAN INDONESIA PERIOD 2015-2016**

Thesis, Faculty of Science and Technology (2021)

(xvi + 111 pages; 42 figures; 23 tables; 5 appendices)

BPJS Kesehatan Indonesia has suffered a deficit for around 13 trillion Rupiah in 2019. With the high deficit rate, BPJS Kesehatan Indonesia has decided to increase the premium around twice of its original price. This step obviously raised a lot of rejections from the society. By utilizing samples data that have been distributed by BPJS Kesehatan Indonesia and the rapid growth of machine learning technology, descriptive analysis and predictive analysis can be conducted to create a solution for this problem. By using K-Means Clustering algorithm, samples can be segmented into clusters. The optimum number of clusters obtained using the elbow method is 4 clusters. The data then clustered into 4 different clusters. This research is also implementing some prediction methods for the number of claims to give stakeholders early warning. All predictions conducted with different models are not satisfactory due to the randomness of the data resulting RMSE (Root Means Square Error) around 0,0113 to 0,00325. This randomness has been confirmed while constructing the ARIMA (AutoRegressive Integrated Moving Average) model. No PACF (Partial Autocorrelation Function) and ACF (Autocorrelation Function) value show high similarities between data lags. As a result of random events, the best model that can predict this phenomenon is the ARIMA model with the order (0,1,0) or the random walk with the drift model where it has the best RMSE value among other models, namely 0,00325. The best conclusion that can be given in this study is to increase the premium to several participants according to their cluster. Cluster 1 is chosen to suffer the increased premium due to its capability and those who are included in cluster 1 mostly are male and comes from productive age segment.

Keywords : BPJS Kesehatan Indonesia, K-Means Clustering, Linear Regression, Random Forest Regressor, ARIMA, Bühlmann-Straub Credibility Theory

Reference : [21] (1987 – 2020)