

# Pengelompokkan Jenis Minyak Mentah Menggunakan Metode X-Means Clustering

Irfani Zuhurfillah

*Information System Department, Harapan Bangsa University  
Raden Patah Street no.100, Central Java, Indonesia*

irfanizuhurfillah@uhb.ac.id

accepted on 26-11-2021

## Abstract

The price of crude oil in Indonesia is still very volatile it cause of strongly influenced by the demand and supply mechanism. The other faktor is the fundamental faktor, the non-fundamental faktor, and the influence and policy of OPEC faktor. Refined petroleum is indispensable in life, used for various human purposes for fuels such as gasoline and kerosene, asphalt, and some of the chemical reagents used to make plastics and medicines. These crude oil prices Information in Indonesia are open and officially published by the Ministry of Energy and Mineral Resources website from 2017 until now. That information is about the monthly prices of several major oil refineries in Indonesia. There is no information about the resume of price fluctuations in one oil refinery, and there is no information on the grouping of oil refineries based on the selling price of the refinery. This study aims to group crude oil price categories in Indonesia using the x-means method. This study is one of the efforts to provide general information about crudes oil price clusters based on major refineries in Indonesia to the public. The results showed that the grouping types of crude oil prices throughout 2020 with the X-Means method contained four optimal clusters with the highest price characteristic being the type of crude oil which is cluster 1, then cluster 0, cluster 3, and the last is cluster 2.

**Keywords:** Metode X-Means, Clustering, Minyak Bumi Indonesia

## I. INTRODUCTION

**M**INYAK bumi merupakan salah satu sumber daya tidak terbarukan yang digunakan oleh manusia secara kontinyu dan berkelanjutan. Sebagai salah satu sumber daya alam yang tidak terbarukan, minyak bumi didapatkan melalui proses studi geologi, analisis sedimen dan berbagai macam studi lainnya [1] [2]. Olahan minyak bumi digunakan untuk berbagai keperluan manusia mulai dari bahan bakar seperti bensin dan minyak tanah, aspal, dan berbagai reagen kimia lainnya yang digunakan untuk membuat plastik maupun obat-obatan [3]. Harga minyak mentah di Indonesia sendiri masih sangat fluktuatif. Fluktuatifnya harga minyak ini sangat dipengaruhi oleh mekanisme permintaan dan penawaran (supply and demand mechanism) [4]. Selain faktor utama tersebut, ada pula beberapa faktor lain yang mempengaruhi harga minyak secara umum yang pertama adalah faktor fundamental, yang terdiri atas permintaan minyak, pasokan minyak, stok minyak, kapasitas produksi cadangan dunia dan kemampuan kilang dunia. Faktor kedua adalah faktor non-fundamental yang

terdiri dari geopolitik, kebijakan pemerintah, cuaca, bencana alam, pemogokan, kerusakan instalasi ranai produksi, pelemahan nilai dolar dan spekulasi. Faktor ketiga adalah pengaruh dan kebijakan OPEC [5]. Naik turunnya harga minyak mentah di Indonesia sebenarnya bersifat terbuka dan diterbitkan secara resmi oleh laman kementerian ESDM mulai dari tahun 2017 sampai sekarang [6]. Akan tetapi informasi tersebut hanya memuat harga setiap bulan dari beberapa kilang minyak utama yang ada di Indonesia dan tidak berisi resume naik turunnya harga di satu kilang minyak dan belum adanya pengelompokan kilang minyak berdasarkan harga jual dari kilang tersebut.

Di sisi lain clustering atau pengelompokan data merupakan metode yang umum digunakan untuk mengkategorikan atau memetakan suatu data berdasarkan kategori tertentu [7]. K-means merupakan salah satu metode clustering yang masih jamak digunakan hingga sekarang. Salah satu penelitian yang menggunakan metode k-means pengelompokan persebaran data kesehatan puskesmas di Indonesia telah dilakukan oleh Burhanuddin dkk, tahun 2016 [8] [1]. Penelitian lainnya yang menggunakan metode k-means adalah penelitian mengenai pengelompokan jenis musik, seperti musik klasik, rap, dan lain-lain. Salah satu kelemahan yang ditemukan di metode k-means adalah waktu komputasi yang lama [9] dan belum dapat menentukan jumlah cluster yang optimal, sehingga jumlah cluster harus dimasukkan secara manual oleh pengguna [10]. Salah satu metode yang dapat mengatasi kekurangan dari metode k-means adalah metode x-means. Metode x-means merupakan pengembangan dari metode k-means sehingga dikenal juga dengan sebutan extended k-means. Metode ini dikembangkan oleh Pelleg & Moore untuk mengatasi kelemahan dalam metode k-means dalam waktu komputasi dan penentuan jumlah cluster (nilai k). Metode x-means dapat menentukan jumlah cluster secara optimal berdasarkan nilai Bayesian Information Criterion (BIC).

Berdasarkan permasalahan dan latar belakang yang telah dipaparkan maka pengelompokan kategori harga minyak mentah di Indonesia dengan menggunakan metode x-means merupakan salah satu upaya yang dapat dilakukan untuk turut serta dalam memberikan gambaran umum kepada masyarakat tentang cluster harga minyak mentah berdasarkan kilang utama yang ada di Indonesia.

## II. LITERATURE REVIEW

### A. X-Means

Algoritma X-Means adalah perbaikan dari metode K-Means yang dikembangkan oleh Dan Pelleg dan Andrew Moore pada tahun 2000, pengembangan algoritma X-Means ditunjukkan untuk mengatasi salah satu kelemahan pada metode K-Means dimana metode K-Means memiliki waktu komputasi yang lama dan jumlah cluster yang harus ditentukan oleh pengguna. Algoritma X-Means menggunakan nilai Bayesian Information Criterion (BIC) dalam menentukan jumlah cluster ideal, sehingga pengguna tidak perlu memasukkan kembali jumlah clusternya, disamping itu metode X-Means juga mengoptimalkan waktu komputasi sehingga lebih cepat dari K-Means [10]. Metode X-means juga digunakan dalam penelitian untuk menentukan jumlah cluster SMK di Jawa Tengah. Dibandingkan dengan K-Means metode X-means mendapatkan nilai DBI sebesar 0,933 dan nilai DBI K-means sebesar 0,914, sedangkan nilai DBI paling besar juga konsisten pada ukuran cluster 10, sebesar 1,439 pada X-means dan 1,322 pada K-means [11]. Penelitian lainnya tentang X-Means juga digunakan untuk mengelompokkan potensi sekolah menengah unggul di Kabupaten Banyumas dimana hasil dari penelitian tersebut menghasilkan nilai cluster 4 dengan nilai DBI sebesar 0,846 [12].

### B. Model Searching

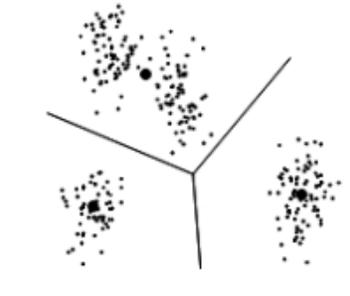
Jumlah cluster pada metode X-Means didasarkan pada pemodelan awal dengan batas bawah (lower bound) jumlah topik (K) yang akan terus bertambah hingga mencapai batas atas (upper bound) dari rentang jumlah K yang telah ditentukan oleh pengguna. Dalam perhitungan diantara interval tersebut jumlah cluster yang memperoleh skor terbaik akan disimpan dan proses eksekusi dihentikan [10].

Operasi yang terus berulang pada metode X-Means sampai berakhir adalah sebagai berikut:

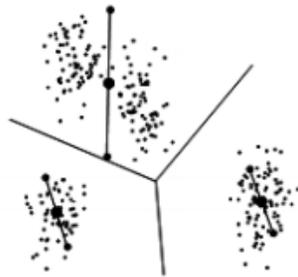
Improve – Params

Improve – Structure

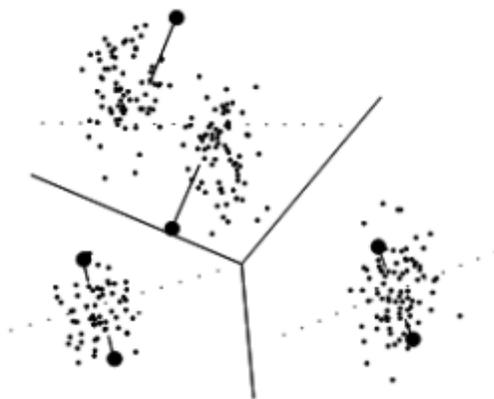
Jika  $k > k_{max}$  maka proses pembentukan cluster akan berhenti dan menghasilkan output jumlah cluster terakhir. Improve-Params adalah algoritma K-Means konvensional yang diulang hingga konvergen (tidak terjadi perpindahan anggota antar cluster). Improve-Structure adalah membagi centroid pada setiap cluster menjadi dua sehingga centroid baru (cluster baru) muncul.



Gambar 1. Hasil klaster dengan  $k = 3$

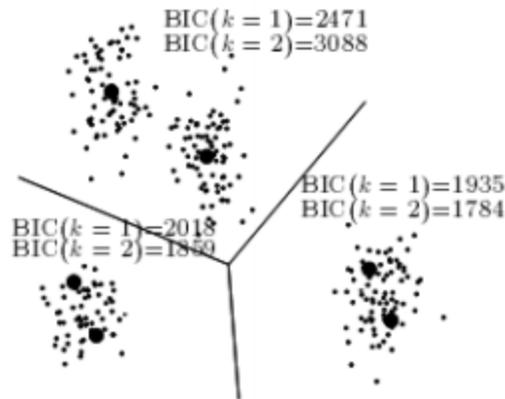


Gambar 2. Pembagian setiap centroid ke dalam 2 children



Gambar 3. Langkah pertama dalam lokal 2 mean (pembagian 2 children baru di setiap centroid)

Gambar 1 menunjukkan pembagian awal cluster menjadi tiga ( $K = 3$ ), dimana hal tersebut merupakan bagian dari proses improve-params atau mengeksekusi data dengan K-Means konvensional dengan nilai  $K = 3$ . Langkah selanjutnya adalah proses improve structure yang ditunjukkan oleh Gambar 2 dan Gambar 3. Proses improve-params dimulai dengan membagi setiap cluster yang telah terbentuk sebelumnya (tiga) menjadi dua centroid baru secara lokal pada setiap cluster (children). Pembagian centroid lokal ini berlawanan arah dalam region cluster lokal searah dengan vector yang dipilih secara acak. Proses selanjutnya adalah mengeksekusi algoritma K-Means pada centroid lokal awal (parent) dengan jumlah cluster dua ( $K = 2$ ) untuk setiap pasang children dengan setiap pasang dibandingkan dengan skor dari region parentnya. Gambar 4 menunjukkan perbandingan skor antara cluster parent dengan cluster children nya (Pelleg and Moore, 2000).



Gambar 4. Hasil setelah eksekusi dari lokal 2 means



Gambar 5. Hasil akhir centroid setelah pengujian lokal

Gambar 5 menunjukkan hasil cluster baru hasil perbandingan skor antara parent dan childrennya dimana ditunjukkan untuk cluster atas nilai  $K = 2$  lebih besar dibandingkan  $K = 1$  sehingga cluster tersebut dibagi

menjadi dua bagian, untuk cluster bagian kanan dan kiri nilai  $K = 1$  lebih besar daripada nilai  $K = 2$  sehingga cluster tetap menjadi satu (tidak terbagi menjadi dua) sehingga hasil yang didapatkan pada iterasi pertama dari awalnya tiga cluster menjadi empat cluster. Proses ini terus berulang sampai mencapai titik konvergen cluster.

### C. Bayesian Information Criterion (BIC) Scoring

Bayesian information criteria (BIC) dalam X-Means digunakan untuk dapat mengetahui model mana yang lebih baik pada saat pembagian dua cluster lokal seperti yang ditunjukkan oleh Gambar 4. Probabilitas posterior digunakan dalam perhitungan BIC dalam memberikan nilai pada model cluster. Formula yang digunakan adalah berdasarkan formula Kass dan Wasserman [13] mulai dari mengukur posterior sampai normalisasi. Formula Kass dan Wasserman ditunjukkan pada formula (1):

$$BIC(M_j) = \hat{l}_j(D) - \frac{P_j}{2} \cdot \log R \quad (1)$$

Dimana

$\hat{l}_j(D)$  adalah fungsi log-likelihood data berdasarkan model ke-  $j$

$P_j$  merupakan jumlah parameter pada  $M_j$ .

Dimana diketahui juga sebagai Schwarz Criterion. Kemudian formula untuk menghitung estimasi probabilitas maksimum atau maximum likelihood estimate (MLE) dengan asumsi gaussian adalah:

$$\hat{\sigma}^2 = \frac{1}{R-K} \sum (x_i - \mu_{(i)})^2 \quad (2)$$

Dengan poin probabilitas:

$$\hat{P}(x_i) = \frac{R_{(i)}}{R} \cdot \frac{1}{\sqrt{2\pi\hat{\sigma}^M}} \exp\left(-\frac{1}{2\hat{\sigma}^M} \left\|x_i - \mu_{(i)}\right\|^2\right) \quad (3)$$

Likelihood data sebagai berikut:

$$l(D) = \log \prod_i P(x_i) = \sum_i \left(\log \frac{1}{2\pi\hat{\sigma}^M} - \frac{1}{2\hat{\sigma}^2} \left\|x_i - \mu_{(i)}\right\|^2 + \log \frac{R_i}{R}\right) \quad (4)$$

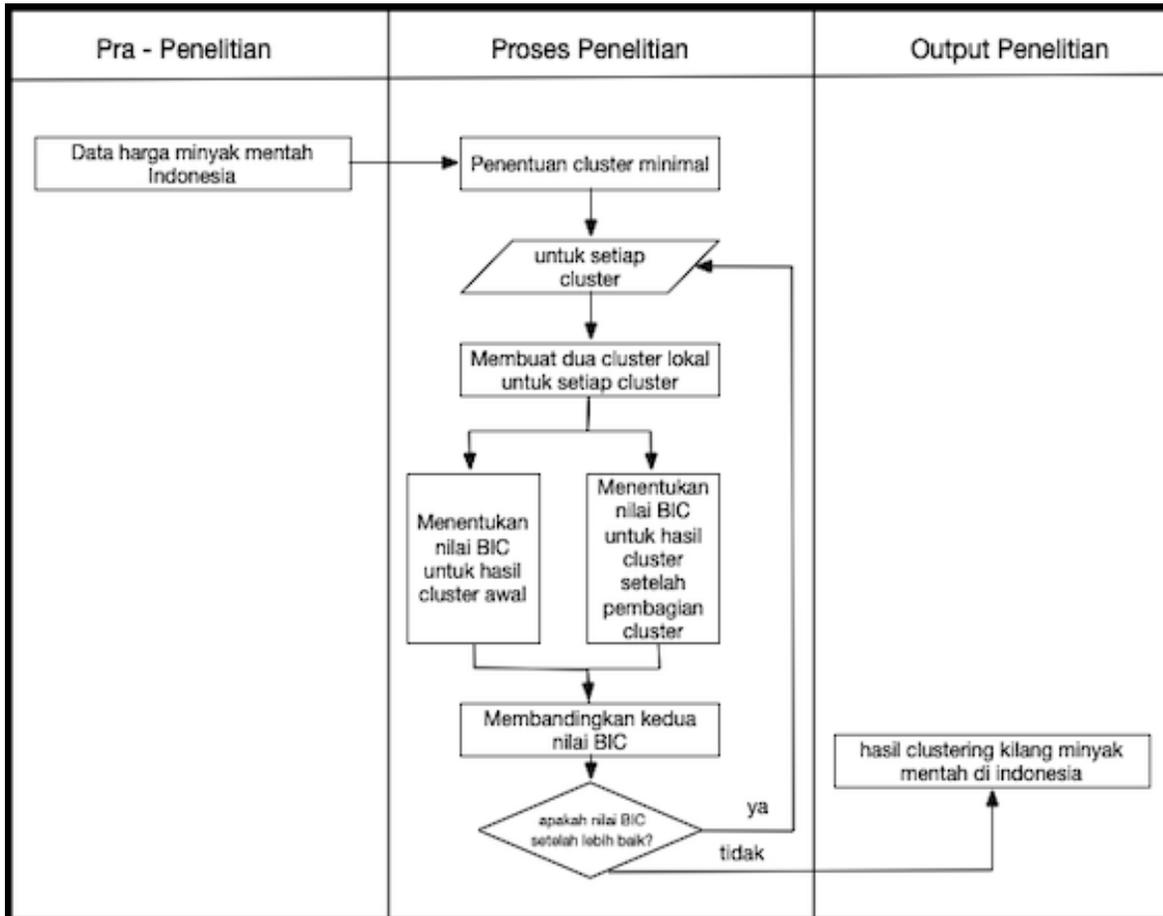
Perhitungan hanya berfokus pada set  $D_n$  dari poin untuk setiap centroid  $n$  dan melakukan perhitungan maksimum likelihood sebagai berikut:

$$\hat{l}(D_n) = -\frac{R_n}{2} \log(2\pi) - \frac{R_n \cdot M}{2} \log(\hat{\sigma}^2) - \frac{R_n - K}{2} + R_n \log R_n - R_n \log R \quad (5)$$

Penggunaan formula BIC secara global digunakan ketika X-means memiliki model terbaik dan digunakan pada pembagian cluster 2 cluster lokal.

### III. RESEARCH METHOD

Metodologi penelitian yang digunakan dalam penelitian ini diawali dengan studi literatur yang dilanjutkan pengumpulan data, analisis dan normalisasi data lalu mengeksekusi data dengan metode X-Means dan terakhir adalah analisis hasil dari pengelompokan data. Desain metode penelitian dapat dilihat pada gambar Gambar 6.



Gambar 6. Desain penelitian

#### A. Pengumpulan data awal

Pengumpulan data merupakan langkah pertama yang dilakukan dalam penelitian ini, pengambilan data minyak mentah yang diambil dari laman kementerian ESDM di laman resmi harga minyak mentah dari kementerian ESDM di Indonesia.

#### B. Analisis dan Normalisasi Data

Kegiatan analisis dan normalisasi data meliputi pemeriksaan data-data anomaly dan data-data kosong yang tidak relevan digunakan dalam penelitian. Tahapan kedua ini merupakan tahapan *pre-processing* data yang akan digunakan, tahapan *pre-processing* dalam penelitian ini berupa pembersihan data yang tidak lengkap, dan atribut data yang kosong.

C. *Pengelompokan data dengan metode x-means*

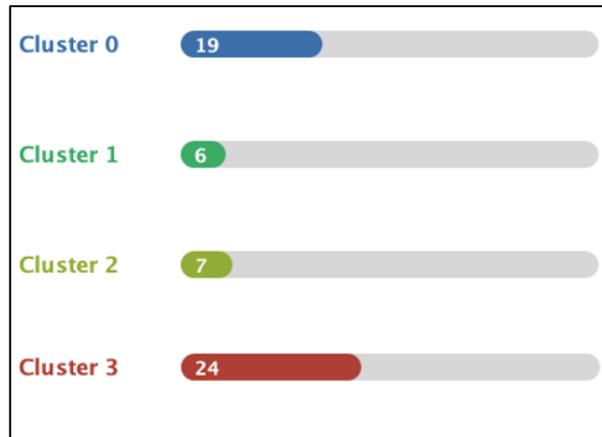
Tahapan selanjutnya setelah dilakukan pembersihan data maka akan langsung diolah menggunakan metode X-Means untuk dapat mengelompokkan data minyak mentah ke dalam jumlah cluster yang konvergen, aplikasi yang digunakan dalam riset ini menggunakan rapidminer untuk melakukan pengelompokan data dengan menggunakan metode X-Means.

D. *Analisis hasil pengelompokan data*

Hasil pengelompokan data menggunakan metode x-means kemudian dianalisis untuk menentukan cluster harga minyak mentah di Indonesia berdasarkan data kilang minyak resmi yang terdaftar di kementerian ESDM.

IV. RESULTS AND DISCUSSION

Data yang diolah pada penelitian ini diambil dari data sepanjang tahun 2020 dengan 7 jenis minyak mentah utama 49 jenis minyak mentah lainnya, pengelompokan dilakukan berdasarkan harga setiap bulan sepanjang tahun sehingga penelitian ini diharapkan dapat mengelompokkan jenis-jenis minyak mentah tersebut sesuai dengan karakteristik harganya di tahun 2020. Dengan menggunakan metode X-Means untuk mengolah data harga jenis minyak mentah tahun 2020 didapatkan sebanyak 4 cluster dengan jumlah anggota sebagai berikut:



Gambar 7. Hasil pembentukan anggota Cluster

Dengan detail setiap anggota cluster dijabarkan pada tabel berikut:

Jenis Minyak Mentah	Cluster
SLC	cluster_0
ARJUNA	cluster_3
ATTAKA	cluster_3
DURI	cluster_1
BELIDA	cluster_3
SENIPAH CONDENSATE	cluster_3
BANYU URIP	cluster_1
ANOA	cluster_3
ARUN CONDENSATE	cluster_3
BEKAPAI	cluster_3
BELANAK	cluster_2
BENTAYAN	cluster_3
BONTANG RETURN CONDENSATE	cluster_3

BULA	cluster_1
BUNYU	cluster_0
CAMAR	cluster_3
CEPU	cluster_2
CINTA	cluster_3
GERAGAI MAKMUR	cluster_0
GERAGAI CONDENSATE	cluster_3
HANDIL MIX	cluster_3
JAMBI	cluster_0
JATIBARANG	cluster_0
JENE	cluster_0
KAJI	cluster_0
KERAPU	cluster_3
KETAPANG	cluster_1
KLAMONO	cluster_1
KOMPLEK PALEMBANG SELATAN	cluster_3
KONDENSAT SAMPANG	cluster_2
KONDENSAT TANGGUH	cluster_2
LALANG	cluster_0
LANGSA	cluster_3
LIRIK	cluster_0
MADURA	cluster_3
MENGOEPEH	cluster_0
MESLU	cluster_2
MUDI MIX	cluster_0
NSC	cluster_3
PANGERUNGAN CONDENSATE	cluster_3
PAMJUATA	cluster_0
PANDAN	cluster_2
PANGKAH	cluster_3
RAMBA	cluster_0
RIMAU	cluster_0
SANGATTA	cluster_0
SELAT PANJANG	cluster_0
SEPINGGAN YAKIN MIX	cluster_3
SOUTH JAMBI CONDENSATE	cluster_3
TANJUNG	cluster_0
TALANG AKAR PENDOPO	cluster_3
TIKA	cluster_1
UDANG	cluster_0
WALIO MIX	cluster_2
WEST SENO BANGKA MIX	cluster_3
WIDURI	cluster_0



Gambar 8. Karakteristik jenis minyak berdasarkan cluster

Berdasarkan Tabel 1 berikut karakteristik dari masing masing cluster tergambar pada Gambar 8. Dimana menunjukkan karakteristik harga pada setiap cluster berbeda, dimana jenis minyak pada cluster 1 merupakan jenis minyak dengan harga rata-rata tertinggi sepanjang tahun disusul dengan jenis minyak pada cluster 0 dan cluster 2. Sedangkan untuk jenis minyak yang masuk pada cluster 3 memiliki karakteristik harga minyak dengan rata-rata terendah sepanjang tahun 2020. Sedangkan untuk keseluruhan cluster rata-rata dari sepanjang tahun 2020 memiliki harga tertinggi di bulan januari dan terendah di bulan April.

## V. Conclusion

Dari pengelompokan jenis harga minyak mentah sepanjang tahun 2020 dengan metode X-Means didapatkan hasil empat cluster optimal dengan karakteristik harga paling tinggi ada pada jenis minyak mentah yang termasuk dalam cluster 1 selanjutnya adalah cluster 0, cluster 3 dan terakhir adalah cluster 2. Saran untuk penelitian selanjutnya adalah menganalisis karakteristik tersebut dengan karakteristik tahun-tahun sebelumnya sehingga dapat diketahui kecenderungan harga jenis minyak mentah tertentu, dan dapat digunakan untuk prediksi karakteristik jenis minyak mentah pada tahun mendatang.

## REFERENCES

- [1] V. Guerriero, S. Vitale, S. Ciarcia and S. Mazolli, "Improved statistical multi-scale analysis of fractured reservoir analogues," *Tectonophysics*, vol. 504, no. 1-4, pp. 14-24, 2011.
- [2] V. Guerriero, A. Lannace, S. Mazolli, M. Parente, S. Vitale and M. Giorgioni, "Quantifying uncertainties in multi-scale studies of fractured reservoir analogues: Implemented statistical analysis of scan line data from carbonate rocks," *Journal of Structural Geology*, vol. 32, no. 9, pp. 1271-1278, 2010.
- [3] S. M. Black, 2006. [Online]. Available: [http://cactus.dixie.edu/sblack/chem1010/lecture\\_notes/2B.htm](http://cactus.dixie.edu/sblack/chem1010/lecture_notes/2B.htm). [Accessed 30 12 2020].
- [4] M. A. Nizar, "Dampak Fluktuasi Harga Minyak Dunia Terhadap Perekonomian Indonesia," *Buletin Ilmiah Litbang Perdagangan*, vol. 6, no. 2, 2012.
- [5] R. A. Fauzannissa, H. Yasmin and D. Ispriyanti, "Peramalan Harga Minyak Mentah Dunia Menggunakan Metode Radial Basis Function Neural Network," *Jurnal Gaussian*, vol. 5, no. 1, pp. 193-202, 2016.
- [6] Kementerian ESDM, 2021. [Online]. Available: <https://migas.esdm.go.id/post/read/harga-minyak-mentah>. [Accessed 20 1 2021].
- [7] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis* 1st Edition, Wiley, 1973.
- [8] A. Burhanuddin, F. M. Wibowo and A. A. Prasetyo, "Pengelompokan Sebaran Tenaga Kesehatan Puskesmas Di Indonesia Dengan Menggunakan Pengklasteran K-Means," in *Konferensi Nasional Sistem Informasi*, Batam, 2016.
- [9] A. Sen, "Automatic Music Clustering using Audio Attributes," *International Journal of Computer Science Engineering (IJCSE)*, pp. 307-312, 2014.

- [10] D. Pelleg and A. Moore, "X-means: Extending K-means with Efficient Estimation of the Number of Clusters," in 17th International Conf. on Machine Learning, Stanford, CA, USA, 2000.
- [11] R. Adhitama, A. Burhanuddin, R. Ananda, "Penentuan Jumlah Cluster Ideal SMK Di Jawa Tengah Dengan Metode X-Means Clustering dan K-Means Clustering Authors", JIKO (Jurnal Informatika dan Komputer), Vol 3, no. 1, pp. 1-5, 2020
- [12] A. Wijayanto, R. Adhitama, "Penggunaan X-Means Clustering Method untuk Mengelompokkan Potensi Sekolah Menengah Unggul di Kabupaten Banyumas", Journal INISTA, Vol. 2, no. 1, pp. 80-88, 2019
- [13] R. E. Kass, L. Wasserman, "A reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion", Journal of The American Statistical Association, Vol. 90, no. 431, pp.928-934, 1995