



EXPLANATORY DATA ANALISIS UNTUK MENGEVALUASI PENELUSURAN KATA KUNCI VIDEO PEMBELAJARAN DI YOUTUBE DENGAN PENDEKATAN MACHINE LEARNING

Mambang Mambang^a, Ahmad Hidayat^b, Finki Dona Marleny^c, Johan Wahyudi^d

^{a,b} Teknologi Informasi, mambang@unism.ac.id, Universitas Sari Mulia

^c Informatika, finkidona@umbjm.ac.id, Universitas Muhammadiyah Banjarmasin

^d Teknik Informatika, johan77@stmik.id, STMIK Indonesia Banjarmasin

ABSTRACT

The purpose of this study was to find correlations related to the variable number of impressions, likes, subscribers, and comments on each learning video keyword search on YouTube. This research uses quantitative methods and experiments with secondary data sources. Exploratory Data Analysis in machine learning using several libraries in Python programming produces image visualizations that provide information related to the dataset that has been processed, such as boxplot graphs, histograms, line plots, and correlation graphs. Exploratory Data Analysis with machine learning that we have done finds results on boxplot graphs on five variables showing a whisker more elongated upwards which states positive data results. The difference in this histogram chart is in the duration variable. On the line plot graph, we find the keywords learning videos learning mathematics have the advantage of four variables and the keywords of accounting learning videos one variable.

Exploratory Data Analysis using the correlation head map in the seaborn library shows that the like and comment variables strongly correlate with a value of 1. Duration variables have a low and negative correlation with other variables. The subscribers variable has a high correlation with the like variable 0.95. Thus, several indicators need to be considered in making learning videos, such as content or content of innovative and creative learning videos, so that the number of likes and comments becomes high. The length of time in learning videos does not affect the number of likes, subscribers, and comments.

Keywords: Explanatory Data Analysis, Keywords, Learning Videos, Machine Learning

ABSTRAK

Tujuan dari penelitian ini untuk menemukan korelasi yang terkait dengan variabel jumlah tayangan, like, subscribers dan komentar pada masing-masing penelusuran kata kunci video pembelajaran di YouTube. Penelitian ini menggunakan metode kuantitatif dan experiment dengan sumber data sekunder. Exploratory Data Analysis pada machine learning dengan menggunakan beberapa library pada pemrograman Python menghasilkan visualisasi gambar yang memberikan informasi terkait dengan dataset yang telah diproses seperti grafik boxplot, histogram, lineplot dan grafik korelasi. Exploratory Data Analysis dengan machine learning yang telah kami lakukan menemukan hasil pada grafik boxplot pada lima variabel menunjukkan whisker lebih memanjang ke atas yang menyatakan hasil data yang positif. Perbedaan pada grafik histogram ini terdapat pada variabel durasi. Pada grafik lineplot kami menemukan kata kunci video pembelajaran belajar matematika memiliki keunggulan empat variabel dan kata kunci video pembelajaran akuntansi satu variabel.

Exploratory Data Analysis dengan menggunakan headmap korelasi pada library seaborn menunjukkan variabel like dan komentar memiliki korelasi yang sangat kuat dengan nilai 1. Variabel durasi memiliki korelasi yang rendah dan negative pada variabel lainnya. Variabel subscribers memiliki korelasi yang tinggi pada variabel like 0.95. Dengan demikian beberapa indikator yang perlu diperhatikan dalam membuat video pembelajaran seperti isi atau konten video pembelajaran yang inovatif dan kreatif, sehingga jumlah like dan komentar menjadi tinggi. Durasi waktu dalam video pembelajaran tidak berpengaruh pada jumlah like, subscribers dan juga komentar.

Kata Kunci: Explanatory Data Analisis, Kata Kunci, Video Pembelajaran, Machine Learning

1. PENDAHULUAN

Peningkatan jumlah data dan informasi dekade ini terus mengalami peningkatan yang sangat signifikan. Visualisasi data dan informasi menjadi sangat penting dalam menunjang kegiatan dalam banyak sektor, terutama data dan informasi pada sektor Pendidikan. Exploratory Data Analysis (EDA) memiliki peran yang sangat penting dalam mengolah data menjadi informasi yang berguna dan memberikan nilai tambah pada penggunaannya. Exploratory Data Analysis (EDA) merupakan salah satu proses dari data science. EDA menjadi sangat penting sebelum melakukan feature engineering dan modeling karena dalam tahap ini harus memahami datanya terlebih dahulu. Video pembelajaran salah satu penunjang dalam proses pembelajaran yang bertujuan mencapai hasil belajar yang ditargetkan[1].

Mengembangkan video pembelajaran memiliki tantangan dalam mempertahankan kualitas isi konten, audio, ide dan kreativitas [2]. Pembelajaran dengan video memberikan pengalaman yang menyenangkan, memotivasi dan menarik bagi penggunaannya [3]. Masalah Pendidikan kini menjadi masalah serius di banyak negara, terutama negara berkembang. Dampak Pandemi yang berlangsung lama turut berdampak pada proses pembelajaran peserta didik. Kondisi inilah memberikan pilihan pada pendidik untuk menggunakan berbagai jenis media pembelajaran yang salah satunya melalui video. Video pembelajaran merupakan suatu media yang dibuat secara sistematis sesuai dengan kurikulum yang berlaku. Video pembelajaran yang banyak terdapat pada platform digital seperti YouTube perlu dilakukan analisis terkait dengan jumlah tayangan, banyaknya like, subscribers dan komentar.

Data dari “We Are Social” sebuah perusahaan yang berfokus dalam melaporkan perkembangan internet, social media dan perkembangan teknologi, menempatkan Indonesia sebagai tiga besar pengguna YouTube terbesar di dunia. Jumlah pengguna YouTube di Indonesia sebanyak 127 juta pengguna. Keterlibatan pengguna dalam mengakses video pembelajaran di Youtube memberikan masukan data pada jumlah penayangan, like, subscribers, durasi dan komentar. Perkembangan platform digital yang sangat cepat memberikan dampak dan perubahan yang signifikan dalam masyarakat [4]. Platform digital memfasilitasi hubungan dan interaksi di masyarakat dengan pengintegrasian sistem [5]. Platform digital juga dapat berfungsi sebagai tempat berbagi donasi dan kegiatan social lainnya [6]. Ekosistem platform digital memberikan keterbukaan akses dan berpusat pada komunitas [7].

Tujuan dari penelitian ini untuk menemukan korelasi yang terkait dengan variabel jumlah tayangan, like, subscribers dan komentar pada masing-masing penelusuran kata kunci. Cara menemukan hasil dari penelitian dengan metode Exploratory Data Analysis (EDA) pada machine learning dengan menggunakan beberapa library pada pemrograman Python.

2. METODOLOGI PENELITIAN

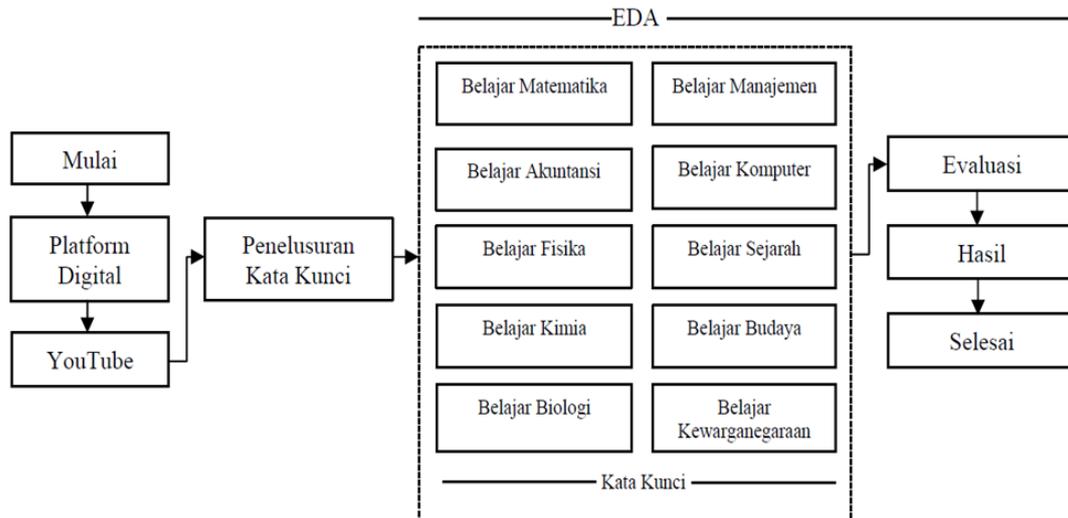
Penelitian ini menggunakan metode kuantitatif dan experiment dengan sumber data sekunder yang terdapat pada Internet pada Platform Digital YouTube. Sumber data yang digunakan pada penelitian diperoleh dengan cara penelusuran kata kunci pada YouTube. Sebagai contoh dengan mengetikkan “Belajar Matematika” pada menu pencarian YouTube. Sebanyak Sepuluh channel yang terkait dengan kata kunci belajar matematika dikumpulkan. Penelusuran kata kunci selanjutnya dilakukan dengan cara yang sama.

Tabel 1. Pengumpulan dataset pada YouTube

Penelusuran Kata Kunci	Penayangan	Like	Subscribers	Durasi	Komentar
Belajar Matematika	18901838	300345	32015100	111	15539
Belajar Akuntansi	2856722	67021	345340	235	3629
Belajar Fisika	2803904	102714	5480640	117	6593
Belajar Kimia	2289870	85086	4944260	137	5041
Belajar Biologi	1807756	52648	5596333	185	3001
Belajar Manajemen	1479753	29483	4326720	104	1355
Belajar Komputer	1199941	28725	76689	132	1545
Belajar Sejarah	1179786	43906	6125000	113	2305

Belajar Budaya	351436	8144	370947	150	674
Belajar	117436	3384	39707	103	540
Kewarganegaraan					

Setelah semua data terkumpul, selanjutnya dilakukan penjumlahan pada masing-masing kata kunci. Hasil dataset yang telah dijumlahkan dapat kami tunjukkan pada tabel 1 diatas.



Gambar 1. Flowchart metode penelitian

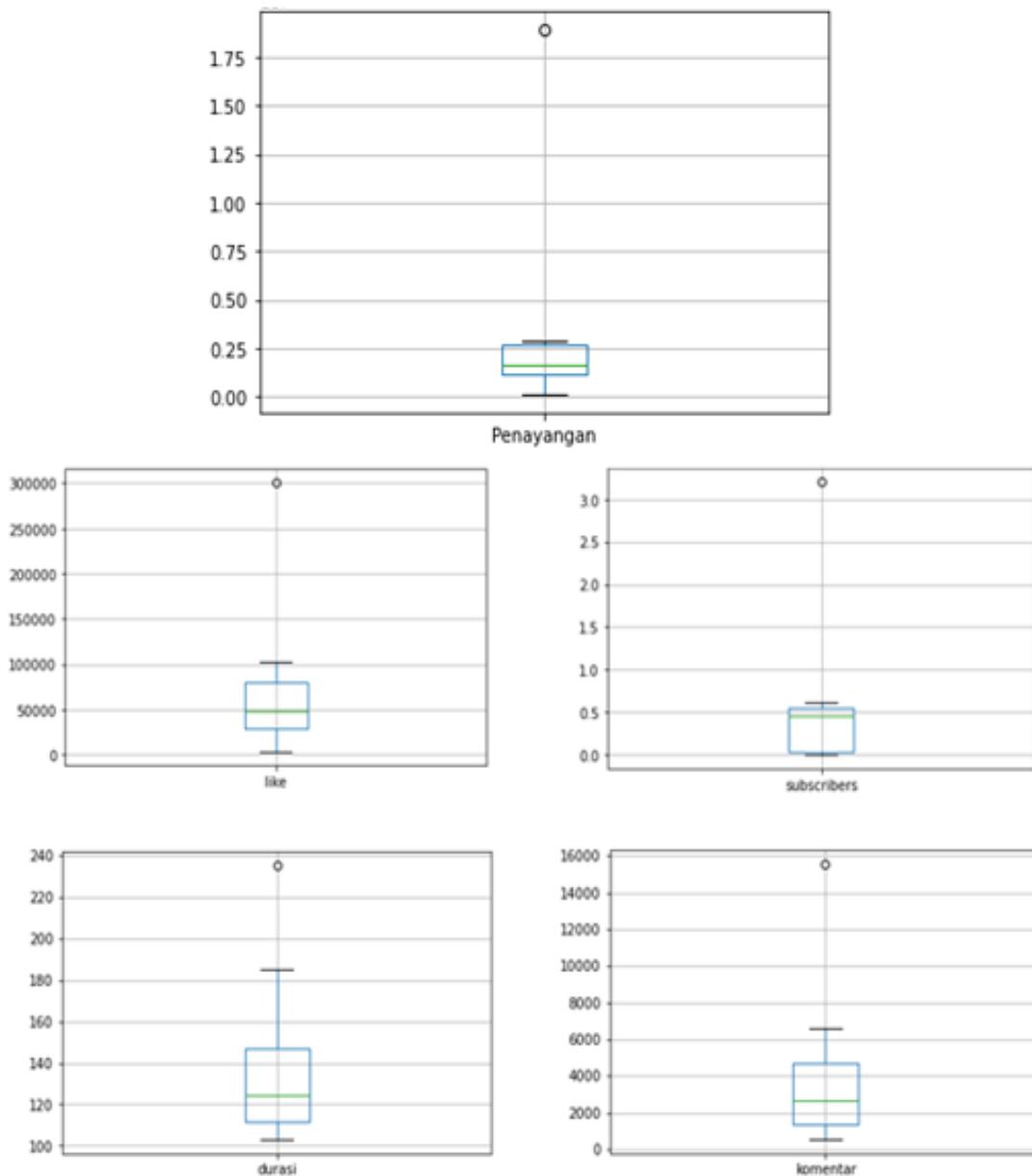
Flowchart pada gambar 1 menunjukkan secara umum tahapan dalam penelitian ini. Data dari tabel 1 selanjutnya dilakukan Exploratory Data Analysis (EDA). Exploratory Data Analysis merupakan bagian dari proses data science [8].

EDA menjadi sangat penting sebelum melakukan feature engineering dan modeling. EDA dapat membantu dalam membangun model yang dapat diverifikasi hasilnya. Analisis data eksploratif dan statistik deskriptif digunakan untuk mengidentifikasi potensi dan pola data yang tidak terekspos secara mendalam. Exploratory Data Analysis memungkinkan analyst memahami isi data yang digunakan, mulai dari distribusi, frekuensi, korelasi dan lainnya.

3. HASIL DAN PEMBAHASAN

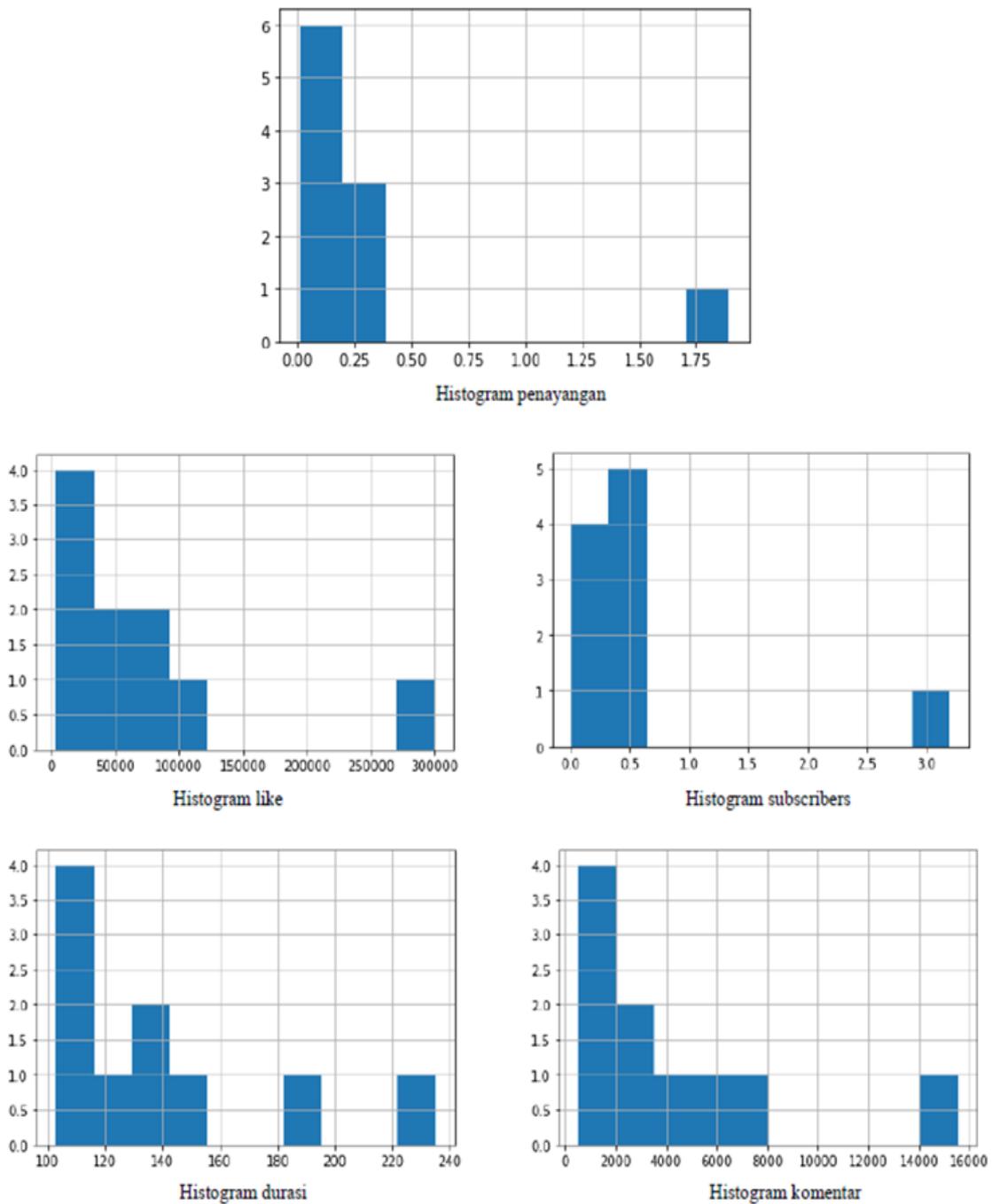
Exploratory Data Analysis pada machine learning dengan menggunakan beberapa library pada pemrograman Python menghasilkan visualisasi gambar yang memberikan informasi terkait dengan dataset yang telah diproses seperti grafik boxplot, histogram, lineplot dan grafik korelasi dengan menggunakan library seaborn.

Exploratory Data Analysis pada penelitian ini dimulai dengan membuat grafik boxplot pada lima variabel seperti penayangan, like, subscribers, durasi dan komentar. Pemanfaatan boxplot sangat berguna dalam membuat ringkasan distribusi sampel yang disajikan secara grafis yang bisa menggambarkan bentuk distribusi data (skewness), ukuran tendensi sentral dan ukuran penyebaran (keragaman) data pengamatan.



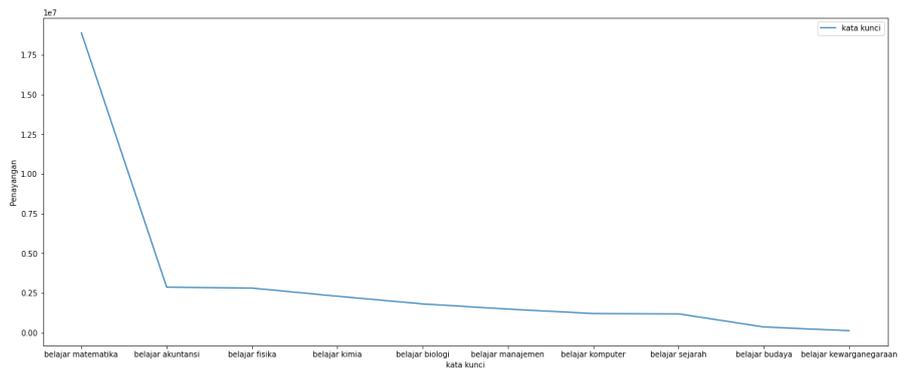
Gambar 2. Boxplot dari lima variabel

Boxplots dapat membantu dalam memahami karakteristik dari distribusi data [9]. Selain untuk melihat derajat penyebaran data (yang dapat dilihat dari tinggi/panjang boxplot) juga dapat digunakan untuk menilai kesimetrisan sebaran data. Panjang kotak menggambarkan tingkat penyebaran atau keragaman data pengamatan, sedangkan letak median dan panjang *whisker* menggambarkan tingkat kesimetrisannya. Adanya *outlier* di bagian atas *boxplot* yang disertai dengan *whisker* bagian atas yang lebih panjang, menunjukkan bahwa distribusi data cenderung menjulur ke arah kanan (*positive skewness*). Sebaliknya, adanya outlier di bagian bawah *boxplot* yang disertai dengan *whisker* bagian bawah yang lebih panjang, menunjukkan bahwa distribusi data cenderung menjulur ke arah kiri (*negatif skewness*). Pada lima variabel *boxplot* diatas menunjukkan *whisker* lebih memanjang ke atas yang menyatakan hasil data yang positif.

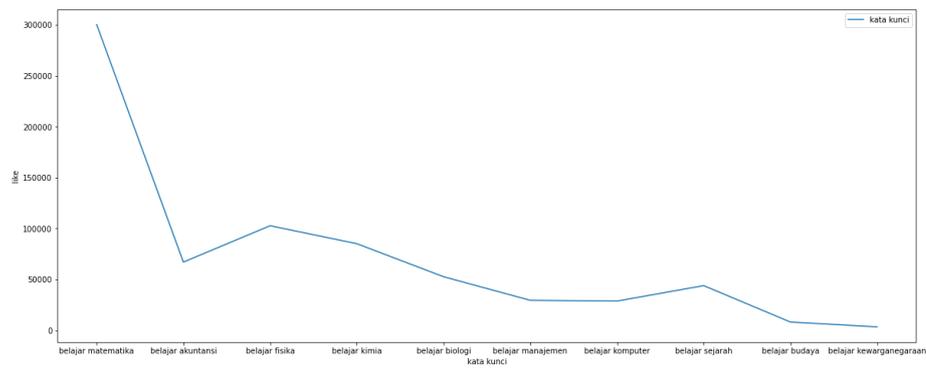


Gambar 3. Histogram dari lima variabel

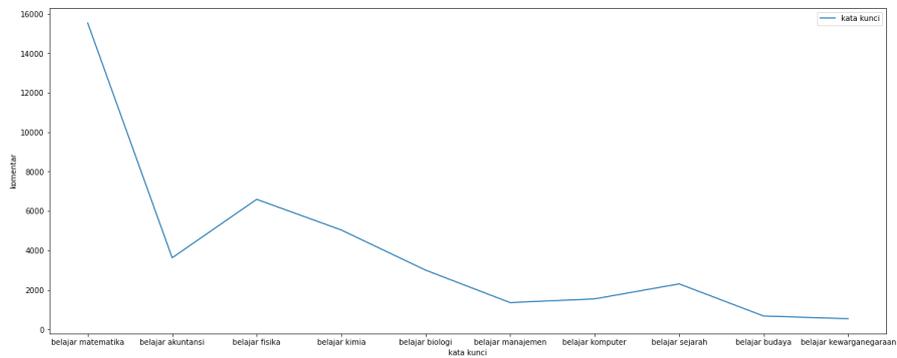
Histogram merupakan tampilan bentuk grafis untuk menunjukkan distribusi data secara visual atau seberapa sering suatu nilai yang berbeda itu terjadi dalam suatu kumpulan data. Distribusi data pada variabel penayangan terbagi dalam 2 bagian, variabel *like* 2 bagian, *subscribers* 2 bagian, *durasi* 3 bagian dan *komentar* 2 bagian. Selanjutnya pada gambar 4,5,6,7,8 merupakan visualisasi grafis dengan menggunakan *lineplot*.



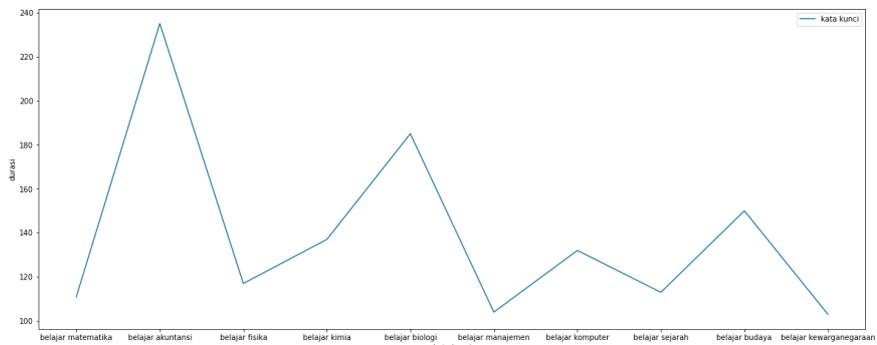
Gambar 4. Lineplot kata kunci dan penayangan



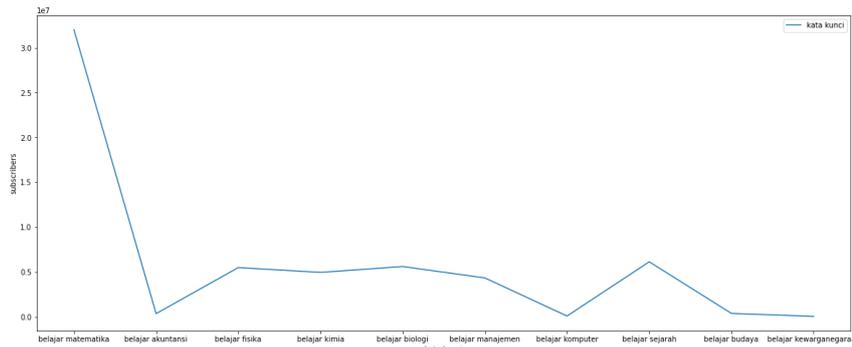
Gambar 5. Lineplot kata kunci dan like



Gambar 6. Lineplot kata kunci dan komentar



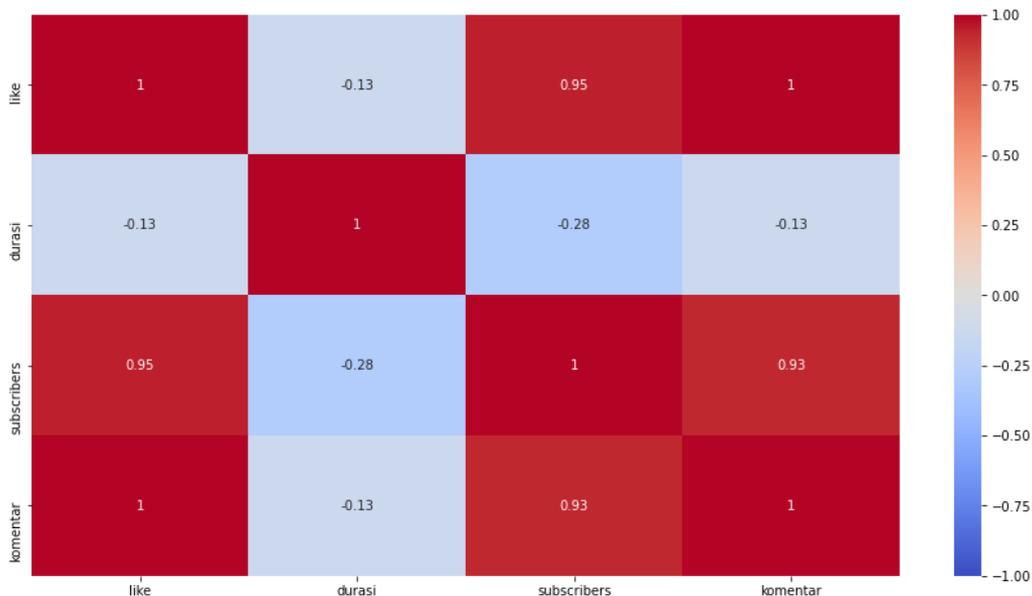
Gambar 7. Lineplot kata kunci dan durasi



Gambar 8. Lineplot kata kunci dan subscribers

Hasil lineplot kata kunci dan penayangan menemukan visualisasi yang menunjukkan video pembelajaran dengan kata kunci belajar matematika memiliki jumlah tayangan, like, komentar dan subscribers yang lebih besar dari kata kunci video pembelajaran lainnya. Sedangkan pada lineplot kata kunci dan durasi kata kunci belajar akuntansi berada paling tinggi dibandingkan dengan variabel lainnya.

Exploratory Data Analysis dengan machine learning yang telah kami lakukan menemukan hasil pada grafik boxplot pada lima variabel menunjukkan whisker lebih memanjang ke atas yang menyatakan hasil data yang positif. Penggunaan dataset pada proses EDA ini dapat dijadikan panduan dan referensi untuk penelitian selanjutnya. Ahmadi dkk, dalam penelitiannya membuat exploratory data analysis faktor eksplorasi terintegrasi dan analisis envelopment data [10]. Pada grafik histogram kami menemukan distribusi data pada variabel penayangan terbagi dalam 2 bagian, variabel like 2 bagian, subscribers 2 bagian, durasi 3 bagian dan komentar 2 bagian. Perbedaan pada grafik histogram ini terdapat pada variabel durasi. Pada grafik lineplot kami menemukan kata kunci video pembelajaran belajar matematika memiliki keunggulan empat variabel dan kata kunci video pembelajaran akuntansi satu variabel. Nikhlis dkk, membuat analisis tingkat kepuasan pengguna google classroom di masa pandemi menggunakan teknik eucs [11]. Analisis dengan berbagai macam metode diperlukan dalam mempersiapkan data dan informasi awal, sehingga dapat menentukan dan pengambilan keputusan selanjutnya.



Gambar 9. Heatmap korelasi variabel dengan library seaborn

Exploratory Data Analysis dengan menggunakan headmap korelasi pada library seaborn menunjukkan variabel like dan komentar memiliki korelasi yang sangat kuat dengan nilai 1. Variabel durasi memiliki

korelasi yang rendah dan negative pada variabel lainnya. Variabel subscribers memiliki korelasi yang tinggi pada variabel like 0.95.

Tabel 2. Validasi hasil korelasi

	like	durasi	subscribers	komentar
like	1.000000	-0.127690	0.952996	0.995742
durasi	-0.127690	1.000000	-0.284623	-0.130522
subscribers	0.952996	-0.284623	1.000000	0.932784
komentar	0.995742	-0.130522	0.932784	1.000000

4. KESIMPULAN DAN SARAN

Exploratory Data Analysis dengan machine learning dengan beberapa library seperti Numpy, Seaborn dan juga Matplotlib menghasilkan visualisasi data yang dapat membantu dan memberikan hasil data yang dapat dijadikan referensi dalam mengambil keputusan dibanyak sektor. Tujuan dari penelitian ini untuk menemukan korelasi yang terkait dengan variabel jumlah tayangan, like, subscribers dan komentar pada masing-masing penelusuran kata kunci. Temuan dari Exploratory Data Analysis dengan penelusuran kata kunci video pembelajaran mendapatkan variabel like dan juga komentar memiliki korelasi yang sangat kuat, sedangkan variabel durasi memiliki korelasi rendah dan negative pada variabel lainnya. Dengan demikian beberapa indikator yang perlu diperhatikan dalam membuat video pembelajaran seperti isi atau konten video pembelajaran yang inovatif dan kreatif, sehingga jumlah like dan komentar menjadi tinggi. Durasi waktu dalam video pembelajaran tidak berpengaruh pada jumlah like, subscribers dan juga komentar.

Ucapan Terima Kasih

Terima kasih yang sebesar-besarnya kepada LPPM Universitas Sari Mulia Banjarmasin yang telah memberikan dukungan baik berupa pendanaan dan juga dukungan pada penyediaan perangkat keras untuk menyelesaikan penelitian ini. Terima kasih juga untuk rekan sejawat yang berada pada Universitas Muhammadiyah Banjarmasin dan STMIK Indonesia Banjarmasin, telah berkontribusi pada penyelesaian penelitian ini.

DAFTAR PUSTAKA

- [1] F. N. Purnomo, "Penggunaan Grup Whatsapp Sebagai Media Pembelajaran di Smp Negeri 1 Berbah Sleman," *Edusaintek J. Pendidikan, Sains dan Teknol.*, vol. 9, no. 2, pp. 430–440, 2022, doi: 10.47668/edusaintek.v9i2.166.
- [2] N. Mohammadhassan, A. Mitrovic, and K. Neshatian, "Investigating the effect of nudges for improving comment quality in active video watching," *Comput. Educ.*, vol. 176, p. 104340, 2022, doi: 10.1016/j.compedu.2021.104340.
- [3] L. Blanquicett, F. Samir, and F. Samir, "Design and Production of Educational Video Games for the Inclusion of Deaf Children," *Procedia Comput. Sci.*, vol. 198, pp. 626–631, 2022, doi: 10.1016/j.procs.2021.12.297.
- [4] V. Ratten, "Digital platforms and transformational entrepreneurship during the COVID-19 crisis," *Int. J. Inf. Manage.*, no. May, p. 102534, 2022, doi: 10.1016/j.ijinfomgt.2022.102534.
- [5] D. Falc, D. Augusto, D. J. Pacheco, and B. Villa, "Moving towards digital platforms revolution? Antecedents, determinants and conceptual framework for offline B2B networks," *J. Bus. Res.*, vol. 142, pp. 344–363, 2022, doi: 10.1016/j.jbusres.2021.12.036.
- [6] E. W. Ho *et al.*, "Findings of living donation experiences shared on a digital storytelling platform: A thematic analysis," *PEC Innov.*, vol. 1, p. 100023, 2022, doi: 10.1016/j.pecinn.2022.100023.
- [7] M. Juutinen, J. Puustinen, and A. Holm, "Digital healthcare platform ecosystem design: A case study of an ecosystem for Parkinson's disease patients," *Technovation*, p. 102551, 2022, doi: 10.1016/j.technovation.2022.102551.
- [8] K. Sorg and H. Khobzi, "A decade of the Swiss electronic vaccination Record: Some insights based on an exploratory data analysis," *Int. J. Med. Inform.*, vol. 158, p. 104660, 2022, doi: 10.1016/j.ijmedinf.2021.104660.
- [9] M. Lancione *et al.*, "Diagnostic accuracy of quantitative susceptibility mapping in multiple system

- atrophy : The impact of echo time and the potential of histogram analysis,” *NeuroImage Clin.*, vol. 34, p. 102989, 2022, doi: 10.1016/j.nicl.2022.102989.
- [10] M. Ahmadi, M. Hassan, M. Osman, and M. Molani, “Integrated exploratory factor analysis and Data Envelopment Analysis to evaluate balanced ambidexterity fostering innovation in manufacturing SMEs,” *Asia Pacific Manag. Rev.*, vol. 25, no. 3, pp. 142–155, 2020, doi: 10.1016/j.apmr.2020.06.003.
- [11] D. Nikhlis, Neilin, Kurniawan, “Analisis Tingkat Kepuasan Pengguna Google Classroom Di Masa Pandemi Menggunakan Teknik Eucs,” *J. Inform. DAN Teknol. Komput.*, vol. 2, no. 1, pp. 1–8, 2022, doi: 10.55606/jitek.v2i1.103.