

# Data Menuju Pengetahuan dengan Pemanfaatan *Data Science*

**Yan Puspitarani\*<sup>1</sup>, Fitrah Rumaisha<sup>2</sup>, Ai Rosita Zakaria<sup>3</sup>  
Sriyani Violina<sup>4</sup>, Azizah Zakiah<sup>5</sup>**

<sup>1,2,3,4,5</sup>Program Studi Teknik Informatika, Universitas Widyatama  
Jl. Cikutra No. 204A Bandung, (022)7275855

e-mail: \*<sup>1</sup>yan.puspitarani@widyatama.ac.id, <sup>2</sup>fitrah.rumaisa@widyatama.ac.id,  
<sup>3</sup>ai.rosita@widyatama.ac.id, <sup>4</sup>sriyani.violina@widyatama.ac.id,  
<sup>5</sup>azizah.zakiah@widyatama.ac.id

## **Abstrak**

*Pengolahan data merupakan hal yang sangat penting bagi institusi pencari laba maupun non laba. Melalui data, institusi mendapatkan pengetahuan yang akan mendukung pihak eksekutif dalam mengambil keputusan untuk strategi bisnisnya. Oleh karena itu, pengolahan data yang baik dapat memberikan dampak yang baik bagi institusi. Institusi bisa mendapatkan berbagai pengetahuan seperti segmentasi pelanggan dan pola pembelian pelanggan untuk kebutuhan marketing, bahkan prediksi kerusakan alat untuk keperluan maintenance dengan memanfaatkan pengolahan data. PT Delima Jaya Katiga membutuhkan pengolahan data untuk melakukan segmentasi dan pola penggunaan jasa pelanggan untuk menemukan potensi pelanggan baru guna menaikkan laba usaha. Oleh karena itu, kegiatan pengabdian ini akan memberikan informasi mengenai data science dan memberikan pelatihan pengolahan data sebagai solusinya.*

**Kata kunci:** *data science, data, pengetahuan*

## 1. PENDAHULUAN

Data menjadi harta karun yang berharga bagi suatu institusi. Pengetahuan yang dihasilkan dari data akan mendukung pihak eksekutif dalam menentukan kebijakan untuk keputusan bisnisnya. Analisis data tidak hanya dilakukan secara matematis dan statistik saja. Tetapi dibutuhkan teknologi yang mumpuni di era *Big Data* ini. Melimpahnya data yang dihasilkan suatu institusi ditambah teknologi yang mumpuni dan pemahaman bisnis yang kuat akan menghasilkan pemodelan data yang lebih akurat. Kemampuan *data scientist* dalam menganalisis data pun diperlukan untuk menghasilkan model yang tepat dari data. Hasil pemodelan ini akan membantu pemangku keputusan untuk memprediksi kemajuan institusi ke depan yang melahirkan kebijakan yang akan diterapkan di institusinya.

Proses pengolahan data inilah yang disebut sebagai *Data science*, dimana pemahaman bisnis, statistika, dan teknologi saling mendukung untuk menghasilkan model data yang akurat dalam mendukung keputusan. *Data science* akan menjawab *business question* yang dimiliki institusi dan hasilnya akan membantu para pemangku keputusan untuk menentukan kebijakan institusi. Pengetahuan yang dihasilkan dari *data science* bisa berupa segmentasi pelanggan dan pola pembelian pelanggan untuk kebutuhan *marketing*, prediksi kerusakan alat untuk keperluan *maintenance*, dan masih banyak hal lainnya.

PT Delima Jaya Katiga (Dejaka) merupakan perusahaan yang menyediakan jasa pemeriksaan K3 (Keselamatan dan Kesehatan Kerja) yang mengeluarkan sertifikasi alat teknik sebagai kelayakan operasional. PT Delima Jaya Katiga melakukan pemeriksaan terhadap peralatan teknik seperti Pesawat angkat/angkut, Pesawat Uap, Bejana Tekan, Tangki Timbun, Elevator, Eskalator, Listrik, dan Instalasi proteksi kebakaran. Pelanggan

yang menggunakan jasa PT Delima Jaya Katiga tersebar di wilayah Indonesia dengan jumlah alat yang diperiksa mencapai puluhan di setiap perusahaan. Banyaknya data yang tersimpan dari hasil pemeriksaan dan adanya peluang untuk menambah pelanggan baru membuat PT Delima Jaya Katiga memerlukan pengolahan data untuk menentukan segmentasi pelanggan dan potensi pelanggan. Selain itu, kepemilikan data hasil pemeriksaan alat juga dapat dimanfaatkan PT Delima Jaya Katiga dalam memprediksi kemungkinan menemukan pola penggunaan jasa oleh perusahaan guna menaikkan laba usaha.

Berdasarkan permasalahan tersebut, pengolahan data menggunakan *data science* sebagai keilmuannya diperlukan untuk membantu PT Delima Jaya Katiga dalam mengembangkan usahanya. Melalui *data science*, data yang dimiliki oleh PT Delima Jaya Katiga akan diolah dan diproses menggunakan bantuan *Machine Learning* untuk mendapatkan segmentasi pelanggan dan pola penggunaan jasa perusahaan. Hasil dari pengolahan data ini akan membantu PT Delima Jaya Katiga dalam menentukan strategi bisnis untuk menambah potensi pelanggan. Kegiatan pengabdian ini akan memberikan informasi kepada mitra dan masyarakat mengenai hirarki DIKW (*Data, Information, Knowledge, Wisdom*), Proses *data science*, kegunaan *data science*, dan beberapa rekomendasi yang diajukan terkait penerapan *data science* pada PT Dejaka sebagai studi kasus.

## 2. METODE PELAKSANAAN

Kegiatan pengabdian ini bertujuan untuk memberikan edukasi kepada mitra dan masyarakat mengenai *data science* dan kegunaannya. Selain itu, diberikan juga beberapa rekomendasi penerapan *data science* pada PT Dejaka selaku mitra kegiatan ini. Oleh karena itu, agar tujuan kegiatan ini tercapai, dilakukan beberapa tahapan kegiatan.

### 2.1 Survei Lapangan

Pada tahap ini dilakukan survei analisis permasalahan dan koordinasi dengan mitra.

### 2.2 Perencanaan

Pada tahap ini dilakukan perencanaan kegiatan berdasarkan hasil survei pada tahap sebelumnya.

### 2.3 Kesepakatan waktu pelaksanaan

Pada tahap ini dilakukan kesepakatan waktu pelaksanaan kegiatan dengan mitra dan masyarakat. Waktu pelaksanaan yang disepakati yaitu Sabtu, 25 Juni 2022 dan dilaksanakan secara *online*.

### 2.4 Implementasi

Pada tahap ini kegiatan pengabdian dilaksanakan secara *online* melalui aplikasi *Zoom meeting*. Pada tahap ini pula, dilakukan analisis kebutuhan mitra terhadap kebutuhan penerapan *data science* di perusahaan.

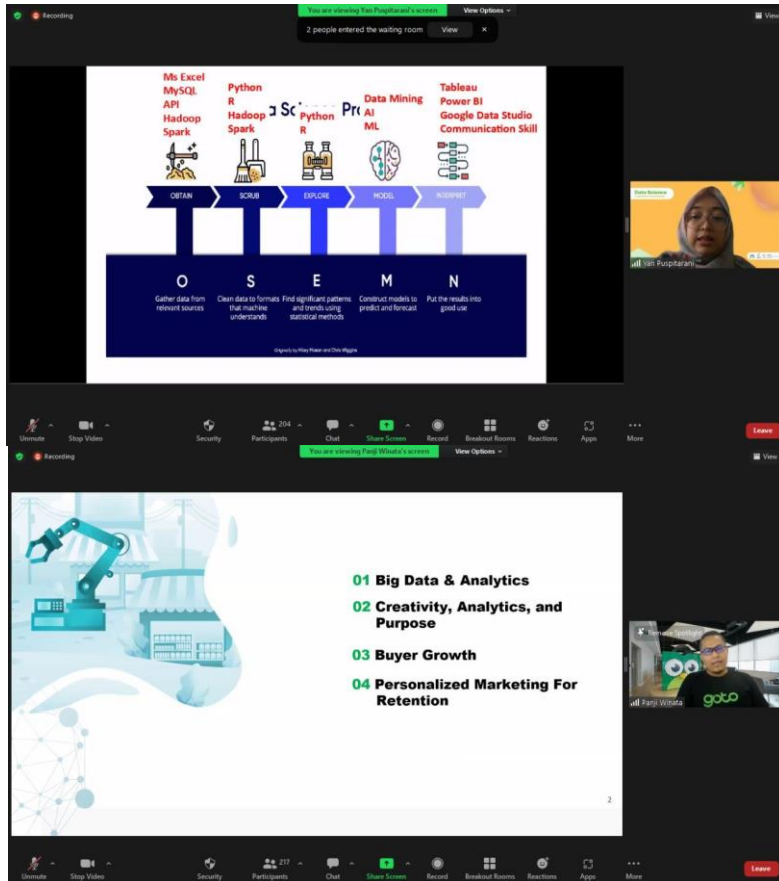
### 2.5 Penyusunan Laporan akhir

Setelah kegiatan dilaksanakan, dilakukan penyusunan laporan kegiatan sebagai bentuk pertanggungjawaban kegiatan pengabdian.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Edukasi webinar “*Data Science : From Data To Knowledge*”

Berdasarkan hasil survey dan koordinasi dengan mitra, ternyata mitra membutuhkan edukasi lebih lanjut mengenai solusi dari permasalahan mereka mengenai pemanfaatan data, yaitu *Data Science*. Agar lebih bermanfaat untuk masyarakat luas, mitra pun setuju untuk mendapatkan edukasi secara *online*.

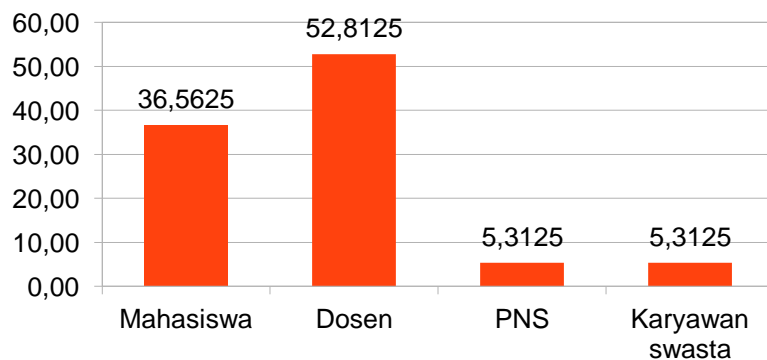


Gambar 1. Kegiatan Pengabdian

Sesuai kesepakatan, kegiatan ini dilaksanakan pada hari Sabtu, 25 Juni 2022 dengan 320 peserta dari kalangan mahasiswa, dosen, PNS, dan karyawan swasta yang tersebar di Indonesia. Berikut ini adalah sebaran peserta webinar berjudul “ *Data Science : From Data To Knowledge*”.

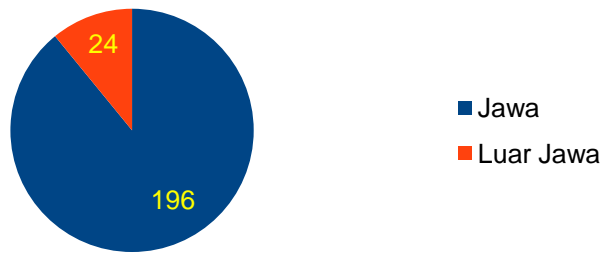
Pada gambar 2 dijelaskan bahwa peserta webinar sebagian besar dari kalangan akademisi, yaitu mahasiswa (36%) dan dosen (52%). Hal ini menunjukkan bahwa peserta dari dosen banyak tertarik dengan tema Data Science. Ketertarikan ini akan menimbulkan minat yang lebih besar terhadap penelitian tentang Data Science dan pembelajaran tentang Data Science.

### Presentasi Sebaran Peserta



Gambar 2. Sebaran Peserta Webinar *Data Science From Data To Knowledge* berdasarkan pekerjaan peserta

### Sebaran Peserta Wilayah Jawa vs Luar Jawa



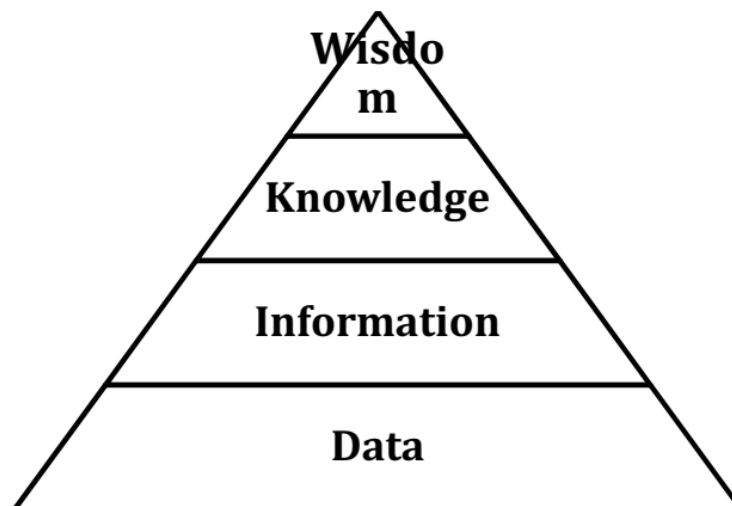
Gambar 3. Sebaran Peserta *Data Science From Data To Knowledge* berdasarkan wilayah

Dari grafik pada gambar 3 di atas dijelaskan bahwa peserta di Pulau Jawa masih lebih banyak dibanding dengan peserta di luar Pulau Jawa. Hal ini menunjukkan bahwa minat peserta terhadap kegiatan ini masih lebih tinggi di Pulau Jawa. Akan tetapi, dengan adanya peserta di Luar Pulau Jawa menunjukkan bahwa sebaran peserta sudah meluas tidak hanya di Pulau Jawa saja. Ini akan menjadi baseline bagi kegiatan berikutnya agar pengetahuan mengenai Data Science dapat diperluas penyebarannya ke Luar Pulau Jawa.

Berikut ini adalah materi yang diberikan dalam webinar *Data Science : From Data To Knowledge*.

#### 1. Hirarki DIKW

Hirarki *Data-Knowledge-Information-Wisdom* (DIKW) digunakan untuk mengidentifikasi dan menggambarkan hubungan antara data, *information*, *knowledge*, dan terkadang *wisdom*. Hubungan ini digambarkan dengan mengidentifikasi entitas di level lebih rendah dalam hirarki, misalnya data ke entitas di level yang lebih atas misalnya *information*. Dari hubungan tersebut dapat diketahui bahwa data dapat digunakan untuk membuat informasi; Informasi dapat digunakan untuk menciptakan pengetahuan, dan pengetahuan dapat digunakan untuk membuat kebijaksanaan. [1]



Gambar 4. Hirarki DIKW

Dalam era Big Data saat ini, pengolahan data menjadi knowledge sangat bermanfaat untuk membantu pihak eksekutif dalam menentukan keputusan dan dijadikan kebijakan (Wisdom). Salah satu ilmu yang mempelajari pengolahan data yang dapat diterapkan untuk menghasilkan knowledge ini adalah *Data science*.

## 2. Proses *Data science*

Strategi bisnis tidak bisa dihasilkan hanya menggunakan feeling saja tapi menggunakan pengalaman. Pengalaman ini dihasilkan dari data-data yang terekam di masa lalu kemudian diolah untuk menghasilkan pengetahuan yang akan menjadi pendukung bagi eksekutif untuk menentukan kebijakan perusahaan. Kemampuan pengolahan data inilah yang disebut sebagai *Data science* dan orang yang melakukannya disebut data scientist.

Seorang data scientist harus mampu menggabungkan ilmu berupa pemrograman, statistik dan matematika untuk mengumpulkan, menafsirkan, dan menganalisis kumpulan data besar yang terstruktur dan tidak terstruktur. Bisa pula dikatakan bahwa *data science* merupakan ilmu yang menggabungkan keahlian di bidang ilmu tertentu dengan keahlian pemrograman, matematika dan statistik. [2] Sehingga, dilihat dari sisi keilmuan, *Data science* dapat diterapkan di berbagai bidang asalkan seorang data scientist memiliki pemahaman bisnis, matematika, statistika, dan teknologi. Sebelum memulai *data science*, yang utama dilakukan adalah menentukan Business Question.

Berikut ini adalah proses *Data science* [3] :



Gambar 5. Proses *Data science*

### 1. *Obtain*

Proses ini melakukan pengambilan data. Untuk melakukan proses ini, diperlukan kemampuan query di database, atau bisa juga data dari Excel dan file text. Skema dan sumber database nya pun bisa berbeda-beda. Selain itu, bisa juga data diambil dari halaman web menggunakan web scrapper atau koneksi menggunakan API. Aplikasi yang paling sering digunakan pada proses ini untuk Big Data adalah Hadoop atau Spark

### 2. *Scrub Data*

Proses ini adalah pembersihan atau *filter* data. Dalam proses scrub ini, data yang tidak relevan akan disingkirkan. Selain itu, dilakukan juga standarisasi format data, mengatur dan merapikan data, membuang apa pun yang tidak diperlukan, dan mengganti data yang hilang serta standarisasi seluruh formatnya. Pada proses ini, biasanya data scientist menggunakan Phyton dan R, tetapi dapat juga menggunakan Hadoop atau Spark jika datanya besar.

### 3. *Explore*

Proses ini akan melakukan penggalian dan pemeriksaan data. Semua data harus diperiksa propertinya, karena tipe data yang berbeda memerlukan perlakuan yang berbeda pula. Kemudian dilakukan pencarian fitur-fitur yang akan digunakan dan mengetes variabel signifikan melalui statistik. Fitur-fitur yang berkorelasi akan digunakan untuk proses sebelumnya. Fitur akan membantu dalam menentukan karakteristik dari masing2 data seperti nama, usia, gender, dan lain-lain. Kemudian, dilakukan visualisasi data untuk melihat trend dan pola dari data menggunakan bar chart atau line chart. Data scientist masih menggunakan Python dan R pada tahap ini.

#### 4. *Model Data*

Pada tahap ini orang lain akan mengatakan “*where the magic happens*”. Pada tahap ini, proses scrub dan explore harus dilakukan dengan benar agar menghasilkan model yang akurat. Tahap pertama pada pemodelan data ini adalah mengurangi dimensi data untuk mempercepat proses pemodelan, kemudian tahap berikutnya dilakukan klasifikasi, clustering, atau forecasting untuk mendapatkan model. Setelah itu, dilakukan test terhadap model yang telah dibuat. Pada tahap ini data scientist memerlukan Machine Learning dan AI menggunakan Python, R, atau software pemodelan data mining seperti Rapidminer, Orange Data Mining dan sebagainya.

#### 5. *Interpreting Data*

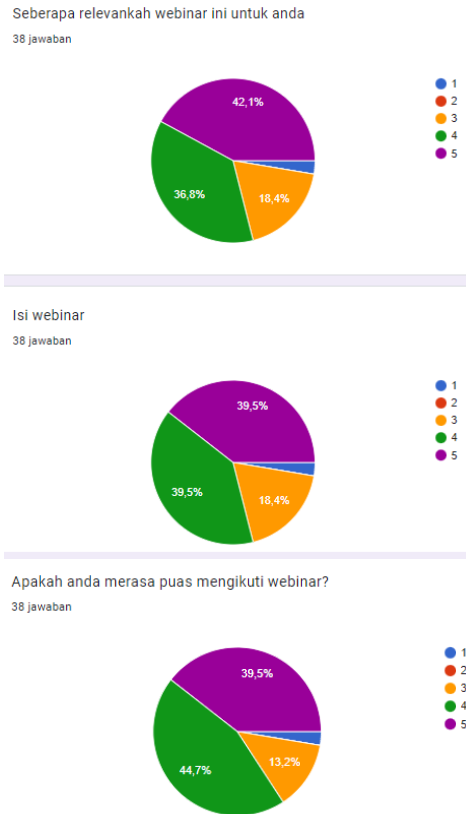
*Output* dari pengolahan data yang diinterpretasi harus bisa dipahami oleh orang-orang awam yang tidak mengerti istilah teknis. Pada tahap ini tim *data science* harus mampu berkomunikasi untuk mempresentasikan hasil temuannya. Presentasi ini bertujuan untuk menjawab persoalan bisnis berdasarkan data yang diperoleh.

### 3. *Kegunaan Data science*

Dalam penerapannya, *data science* digunakan dalam hal-hal berikut ini.

1. E-commerce  
Penjual online secara otomatis akan menyesuaikan etalasenya berdasarkan profil data pembeli. Beberapa online shop juga menyesuaikan harga berdasarkan profil penghasilan konsumen, atau disebut dengan harga yang dipersonalisasi [4].
2. Fraud detection  
Kegiatan anomali dari transaksi finansial dari perusahaan jasa finansial [5].
3. Dynamic Pricing  
Pelanggan dilakukan segmentasi kemudian Setiap kelompok pelanggan dapat ditawari produk dengan harga berbeda [6].
4. Healthcare  
Pada bidang kesehatan, pemanfaatan data berupa gambar dimanfaatkan untuk deteksi penyakit komorbid atau tingkat keganasan kanker dan masih banyak lagi [7].
5. Mesin pencarian di internet  
Orang-orang yang melakukan pencarian di internet mendapatkan rekomendasi kata kunci dari kata pertama yang diketik di mesin pencari. Hasil pencariannya pun biasanya dikelompokkan berdasarkan konten hasil pencarian
6. Targeted Advertising  
Iklan produk akan muncul di halaman website atau media sosial yang sedang dibuka berdasarkan riwayat pencarian pengguna [8]. Hal ini lebih bermanfaat karena iklan yang ditampilkan sesuai dengan kebutuhan penggunanya.
7. Website Recommendation  
Website akan menampilkan rekomendasi produk secara otomatis berdasarkan kebutuhan pengguna sekaligus menampilkan perbandingan spesifikasi masing-masing produk yang direkomendasikan [9].
8. Advanced Image Recognition  
Wajah orang dalam foto di media sosial dapat dikenali dan di-tag secara otomatis. Selain itu, pencarian foto menggunakan sampel wajah juga dapat dilakukan menggunakan Image Recognition [10].

Setelah pelaksanaan webinar, para peserta diminta untuk mengisi kuisisioner mengenai kegiatan webinar. Kuisisioner ini diberikan kepada 38 peserta saja sebagai sampel untuk mengisi form yang disebar. Hasil dari kuisisioner digambarkan pada Gambar 6 berikut.



Gambar 6. Hasil Kuisisioner Peserta

Berdasarkan hasil kuisisioner, terlihat bahwa dari skala 1 (buruk) sampai 5 (baik sekali), sebagian besar memberikan penilaian baik. Hal ini berarti, kegiatan webinar ini memuaskan para peserta.

### 3.2 Rekomendasi Penerapan Data science Pada PT Dejaka

Dashboard Business Intelligence berupa rekapitulasi jumlah pelanggan/mitra, alat, status pengerjaan, pengeluaran operasional, dan pendapatan dapat dilakukan untuk memudahkan pemilik usaha melihat kemajuan kegiatan perusahaannya. Setelah didapatkan hasil rekapitulasi, *data science* dapat diterapkan untuk melakukan forecasting terhadap jumlah alat yang akan diperiksa pada periode mendatang. Forecasting yang dilakukan dapat memanfaatkan berbagai algoritma Statistika atau Machine Learning.

Selain itu, dengan memanfaatkan data mining terhadap frekuensi mitra yang melakukan kerjasama, dapat dihasilkan pola pemesanan mitra untuk pemeriksaan alat dapat diprediksi sehingga pemilik perusahaan dapat menentukan perencanaan untuk operasional masa depan. Untuk mendapatkan pola pemesanan, berbagai algoritma Association Analysis dalam Data Mining dapat diterapkan.

Dengan memanfaatkan *data science* juga, PT Dejaka dapat melakukan segmentasi pelanggan dan potensi pelanggan baru dilihat dari jumlah kebutuhan pemeriksaan alat di setiap daerah. Hal ini, berguna untuk kebutuhan marketing dalam memasarkan jasa PT Dejaka. Untuk melakukan segmentasi pelanggan, algoritma-algoritma Clustering dapat dilakukan.

#### 4. KESIMPULAN

Berdasarkan tahapan analisis, akhirnya kegiatan pengabdian ini menghasilkan dua luaran, yaitu edukasi mengenai *data Science* dan rekomendasi bagi mitra mengenai penerapan *data Science* yang mungkin dibangun pada PT Dejaka.

Melalui edukasi, para peserta dan mitra mendapatkan pengetahuan tentang pentingnya *data science* dalam berbagai bidang di dunia industri terutama bidang marketing. Para peserta mengetahui bahwa data merupakan aset yang sangat berguna jika diolah dan menghasilkan pengetahuan untuk membantu dalam pengambilan keputusan. Sedangkan bagi mitra, rekomendasi yang telah diajukan akan sangat membantu dalam proses bisnis marketing di PT Dejaka.

Kegiatan webinar dapat dikatakan berhasil karena sebaran pesertanya yang mencapai luar jawa dengan jumlah peserta mencapai 320 peserta dari berbagai kalangan. Selain itu, hasil kuisisioner yang diberikan kepada beberapa peserta sebagai sampel menyatakan bahwa mereka puas dengan kegiatan webinar ini.

#### 5. SARAN

Perlunya kegiatan praktek *Data science* pada kegiatan PKM mendatang agar masyarakat lebih memahami proses yang terjadi dalam *Data science*. Dengan melakukan praktek, masyarakat akan lebih menghargai data yang disimpan sebagai aset yang berharga dalam pengambilan keputusan.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada LP2M Universitas Widyatama yang telah memberi dukungan *financial* terhadap kegiatan PkM ini dengan nomor kontrak No. 002/SPC2/LP2M-UTAMA/VI/2022.

#### DAFTAR PUSTAKA

- [1] Jennifer Rowley, "The wisdom hierarchy: representations of the DIKW hierarchy," *Journal of Information and Communication Science*, 2007.
- [2] Data Robot. [Online]. <https://www.datarobot.com/wiki/data-science/>
- [3] Cher Han Lau. (2019, January) Towards *Data science*. [Online]. <https://towardsdatascience.com/5-steps-of-a-data-science-project-lifecycle-26c50372b492>
- [4] Hussain Saleem, Altaf Hussain Nizamani, Samina Saleem, and Jamshed Butt, "DATA SCIENCE AND MACHINE LEARNING APPROACH TO IMPROVE E-COMMERCE SALES PERFORMANCE ON SOCIAL WEB," *International Journal of Advanced Research in Engineering and Technology (IJARET)*, vol. 12, no. 4, pp. 401-424, April 2021.
- [5] Nadya Intan Mustika, Bagus Nenda, and Dona Ramadhan, "Machine Learning Algorithms in Fraud Detection: Case Study on Retail Consumer Financing Company," *Asia Pacific Fraud Journal*, vol. 6, no. 2, pp. 213-221, December 2021.

- [6] Mar'atus Sholikhhatun Nisa and Yusuf Amrozi, "Pemanfaatan Big Data Dalam Penerapan Dynamic Pricing (Studi Kasus Amazon.Com) ," *JSM STMIK Mikroskil* , vol. 20, no. 1, pp. 33-40, April 2019.
- [7] Chandan K. Reddy and C. Aggarwal Charu,,: CRC Press, 2015.
- [8] Araoye Oluwasegun A. (2019, September) Medium Web Site. [Online]. <https://medium.com/@araoyeoluwasegunvhictor/the-application-of-data-science-in-targeted-advertising-e81c1f7f5b62>
- [9] Vishal Shah and Shridevi S, "Data-Science Recommendation System using Semantic Technology," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 9, no. 1, pp. 2592-2599, October 2019.
- [10] Parvathy. (2020, January) Free Press Journal. [Online]. <https://www.freepressjournal.in/sponsored-content/how-data-science-is-helpful-in-advanced-image-recognition>