

## Basic class soal Data Science bagi Startup



Bisnis startup memang selalu berhubungan dengan data. Kebanyakan startup telah memiliki saluran data yang baik. Akan tetapi, tidak sedikit yang masih belum memiliki proses analisis yang dapat direproduksi. Padahal, aspek data science ini sangat penting untuk mendorong perkembangan bisnis.

Kali ini kita akan mempelajari dasar-dasar mengenai data science dalam bisnis startup. Selain para analis data, Anda juga bisa mempelajari materi ini agar nantinya lebih mudah ketika bekerja sama dengan para analis. Dengan mempelajari dasar-dasar data science Anda akan paham bagaimana para ilmuwan bekerja menjalankan eksperimen dan membangun produk data. Sebagai catatan, minimal Anda harus paham mengenai bahasa pemrograman Java atau R.

### **Kenapa harus mempelajari data science?**

Mungkin sebagian dari Anda mulai bertanya-tanya, bagaimana data science bisa meningkatkan produk? Secara langsung memang tidak demikian. Namun, dengan data science Anda dapat mengumpulkan dan mempelajari pola perilaku pelanggan. Dengan demikian Anda bisa meningkatkan kualitas produk berdasarkan data yang diperoleh.



### Data Scientists

#### Get more done, faster, on more data

- Start coding faster – no setup required
- Use existing Python and R libraries
- Access secure Hadoop environments
- Easily automate, share, and deploy project code



### Analytics Leaders

#### Scale the data science practice

- Deliver more projects, faster
- Easily onboard team members
- Enforce library standards, drive reuse
- Enable collaboration between data scientists and business teams.



### IT/Operations

#### Reduce costs, risk, and fragmentation

- Deliver self-service, shared analytics infrastructure
- Leverage existing Hadoop investments
- Maintain security, governance, auditability, and SLAs

Selain itu, beberapa manfaat ini akan Anda peroleh berkat menggunakan data science:

- Mengidentifikasi metrik bisnis utama untuk melacak dan memperkirakan strategi
- Membangun model prediksi perilaku pelanggan
- Menjalankan eksperimen untuk menguji perubahan produk
- Membangun produk data yang memungkinkan fitur produk baru

Rata-rata bisnis terjebak pada dua atau tiga langkah awal, kemudian tidak memanfaatkan potensi penuh data science. Tapi percayalah, beberapa pengetahuan dasar terkait data science berikut ini sangat bermanfaat bagi bisnis Anda.

### Ikhtisar data science

Berikut ini dasar-dasar dalam data science yang perlu Anda pelajari:

- **Tracking Data:** Bagaimana cara mengambil data dari aplikasi dan halaman web, mencoba metode yang berbeda untuk melakukan tracking data, meningkatkan fokus terhadap masalah privasi dan penipuan, hingga memanfaatkan tool Google PubSub.
- **Data pipelines:** Menyajikan berbagai pendekatan untuk mengumpulkan data yang digunakan oleh tim analisis dan data science, membahas pendekatan

dengan flat file, basis data, dan data lakes, hingga menyajikan implementasi menggunakan PubSub, DataFlow, dan BigQuery.

- **Business Intelligence:** Mengidentifikasi praktik umum untuk ETL, laporan/ dasbor otomatis, hingga menghitung metrik dan KPI run-the-bisnis. Langkah ini memanfaatkan beberapa tool, seperti R Shiny, Apache Airflow, dan Data Studio.
- **Exploratory Analysis:** Meliputi analisis umum yang digunakan untuk menggali data seperti membangun histogram, fungsi distribusi kumulatif, analisis korelasi, dan fitur penting untuk model linier. Menyajikan analisis contoh dengan set data publik Natalty.
- **Predictive Modeling:** Membahas pendekatan untuk pembelajaran yang diawasi dan tidak diawasi, menyajikan model prediktif churn dan lintas promosi, hingga metode untuk mengevaluasi kinerja model offline.
- **Model Production:** Menunjukkan cara meningkatkan model offline untuk mencetak jutaan rekaman, dan membahas pendekatan batch dan online untuk penyebaran model. Hal ini meliputi Productizing Data Science di Twitch, dan Model Producizting dengan DataFlow.

# Benefits Of Test Marketing

- Expose product to marketing environment and assess sales performance
- Identify weaknesses in product or marketing mix
- Experiment with variations in marketing mix
- Reduce risk of failure

57

- **Eksperimen:** Mempelajari masalah A/B Testing, membahas cara menyiapkan kerangka eksperimen untuk menjalankan eksperimen, dan menyajikan analisis contoh dengan R dan bootstrapping.
- **Recommendation Systems:** Mempelajari dasar-dasar sistem rekomendasi dan bagaimana cara meningkatkan rekomendasi untuk sistem produksi.
- **Deep Learning:** Mempelajari pengenalan ringan untuk masalah data science terbaik dan ditangani dengan pembelajaran mendalam, seperti menandai pesan obrolan sebagai ofensif.

## Berbagai tool dalam data science

Dalam data science Anda dapat memanfaatkan tool Google Cloud Platform (GCP). Kelebihan GCP salah satunya yakni menyediakan sejumlah layanan terkelola yang memungkinkan tim kecil membangun jaringan data pipeline, menghasilkan model prediktif, dan memanfaatkan deep learning.

## Benefits of using GCP

True cloud economics	Built for scale	Get access to innovation	Google-grade security
<p>Make the cloud work for you.</p> <ul style="list-style-type: none"><li>• Per-minute pricing</li><li>• Sustained use discounts</li><li>• Custom machine types, exact right infra at the right cost</li><li>• Tailor-made offerings priced to match use case (e.g. Nearline, Preemptible VMs)</li></ul>	<p>Scale your business smoothly and responsibly.</p> <ul style="list-style-type: none"><li>• Live migration</li><li>• Private network, largest ISP, owned fibre</li><li>• Carbon-neutral operation</li><li>• \$10B/Year investment, 2X competitors</li></ul>	<p>15 years of support for Open Source software</p> <ul style="list-style-type: none"><li>• Google's track record of open innovation continues- allows you to deploy at your own rate:<ul style="list-style-type: none"><li>◦ Container management</li><li>◦ Machine learning</li><li>◦ Scalable databases</li><li>◦ No-ops SW development</li></ul></li></ul>	<p>The best security available, because we need it.</p> <ul style="list-style-type: none"><li>• Ongoing investment and innovation in physical and electronic countermeasures</li><li>• Over 500 security engineers building technology.</li><li>• Google controls its cloud stack, from silicon up.</li></ul>

Untuk bahasa pemrograman, saya rekomendasikan menggunakan R untuk scripting, Java untuk produksi, serta SQL untuk bekerja dengan data di BigQuery. Anda juga dapat menggunakan tool lain seperti AirFlow dan Shiny. Keep learning!

Source:

<https://towardsdatascience.com/data-science-for-startups-introduction-80d022a18aec>  
<https://www.datascience.us/getting-started-cloudera-data-science-workbench/>  
<https://www.slideshare.net/chegala5626/chp-3-the-business-of-product-management>  
<https://www.slideshare.net/DataStax/cassandra-on-google-cloud-platform-ravi-madasu-google-ben-lackey-datastax-c-summit-2016>  
<https://www.pexels.com/>