



## Ujian Modul 3

---

Purwadhika Digital Technology School

JCDSA - 0402 (BSD) & JCDSA - 0404 (JKT)

---

- **Baca soal baik-baik. Jika ada pertanyaan mengenai soal, silakan ditanyakan.**
  - **Ujian ini bersifat *open book*. Kalian boleh melihat catatan atau mencari materi di internet, tapi mohon untuk tidak berdiskusi satu sama lain.**
  - **Peserta diwajibkan menyalakan webcam dan menonaktifkan microphone selama ujian berlangsung.**
  - **Kalian memiliki waktu 4 jam untuk mengerjakan ujian ini.**
  - **Peserta diperkenankan untuk menuliskan asumsi atas jawaban yang diberikan jika terdapat batasan-batasan yang tidak dijelaskan oleh soal.**
  - **Periksa kembali kelengkapan dan kesesuaian jawaban Anda. Peserta hanya diberi kesempatan untuk mengirimkan dokumen jawaban ke email pengajar sebanyak maksimal dua kali.**
  - **Peserta yang telah mengumpulkan dokumen jawaban dipersilakan untuk meninggalkan Zoom Meeting.**
- 

### Case Study

Di dalam bisnis perhotelan, memastikan ketersediaan kamar hotel di suatu waktu dan untuk beberapa waktu yang akan datang sangat penting dalam mengoptimalkan pendapatan. Informasi mengenai ketersediaan kamar yang akurat dan aktual dapat menghindarkan hotel dari berbagai kerugian, semisal kerugian yang disebabkan oleh tamu yang membatalkan pesanan atau tidak melakukan check in di waktu kedatangannya.

Tamu yang membatalkan penginapannya, baik setelah mengabari pihak hotel terlebih dahulu (misalkan menghubungi sendiri atau dihubungi pihak hotel) maupun tidak, menyebabkan kekosongan kamar yang seharusnya mereka inapi. Kamar-kamar ini sebenarnya masih dapat disewakan kembali ke tamu yang lainnya, namun akan lebih sulit untuk memutuskan dengan segera kamar mana saja yang dapat disewakan kembali apabila kabar pembatalan oleh tamu diterima pihak hotel secara mendadak, semisal beberapa saat sebelum waktu check in.

Tamu yang membatalkan penginapannya, baik setelah mengabari pihak hotel terlebih dahulu (misalkan menghubungi sendiri atau dihubungi pihak hotel) maupun tidak, menyebabkan kekosongan kamar yang seharusnya mereka inapi. Kamar-kamar ini sebenarnya masih dapat disewakan kembali ke tamu yang lainnya, namun akan lebih sulit untuk memutuskan dengan segera kamar mana saja yang dapat disewakan kembali apabila kabar pembatalan oleh tamu diterima pihak hotel secara mendadak, semisal beberapa saat sebelum waktu check in.

Diberikan sebuah dataset berisi informasi pemesanan kamar hotel, Anda diminta untuk membangun sebuah model untuk memprediksi apakah seorang pelanggan yang telah memesan sebuah kamar akan membatalkan pesannya. Model Anda diharapkan dapat menjadi salah satu sarana untuk mengidentifikasi pembatalan pesanan dengan lebih dini sehingga dapat mengurangi risiko kerugian akibat pembatalan pesanan seperti yang dijelaskan di atas.

---

## Dataset

Dataset ini berisi informasi mengenai pemesanan tamu sebuah resort hotel dan city hotel. Atribut yang tersedia di dalamnya meliputi tanggal menginap, selisih hari antara tanggal pemesanan dengan tanggal menginap, durasi menginap, jumlah tamu, status pemesanan, dan keterangan-keterangan lainnya berkaitan dengan suatu pemesanan maupun pelanggan yang membuat pesanan itu sendiri.

Anda dapat mengakses dataset tersebut [disini](#)

---

Anda perlu menjawab kasus di atas dengan mengikuti instruksi di bawah ini. Tuliskan jawaban Anda di dalam sebuah Jupyter Notebook. Sertai penjelasan untuk setiap bagian kode dalam bentuk comment dan untuk tahap analisis Anda dalam bentuk markdown. Cantumkan referensi apabila Anda mengutip penjelasan dari sumber luar.

### 1. Problem Framing (10 poin)

- *Business objective* apa yang hendak dicapai melalui proyek ini?
- Keluaran (output) seperti apa yang akan dihasilkan dari proyek ini dan bagaimana end user atau stakeholder akan memanfaatkan keluaran tersebut?
- Machine learning task apa yang akan dikerjakan? (supervised/unsupervised learning, classification/regression/clustering/dimensionality reduction problem dsb.)
- Performance measure apa yang akan digunakan di dalam proyek ini?
- Risiko seperti apa yang mungkin diakibatkan oleh kesalahan prediksi dari model Anda?

### 2. EDA (20 poin)

Lakukan eksplorasi terhadap dataset tersebut untuk memperoleh gambaran mengenai kondisi nilai di masing-masing variabel yang mungkin untuk dimanfaatkan untuk prediksi, variabel mana saja yang butuh untuk diolah, serta mengidentifikasi variabel mana saja yang akan digunakan untuk melakukan prediksi. Tuliskan penjelasan untuk setiap temuan.

### 3. Data Preparation (20 poin)

Lakukan data cleaning, feature selection, dan feature engineering berdasarkan temuan-temuan pada tahap sebelumnya. Jelaskan argumentasi Anda untuk setiap langkah data preparation yang diambil.

#### 4. Model Training (25 poin)

Pilihlah paling banyak tiga kandidat model dan lakukan training untuk masing-masing model. Jelaskan pula alasan pemilihan kandidat model maupun langkah-langkah yang diterapkan dalam melakukan training.

#### 5. Evaluation, Model Selection, and Model Tuning (15 poin)

Evaluasi masing-masing model yang telah Anda latih, kemudian pilih yang menurut Anda memiliki kinerja terbaik, lalu lakukan tuning terhadapnya. Setelahnya, lakukan evaluasi akhir. Ulangi langkah-langkah sebelumnya bila perlu.

#### 6. Result, Conclusion, and Recommendation (10 poin)

Tuliskan kesimpulan dari proyek ini. Hubungkan dengan studi kasus di awal serta berikan rekomendasi untuk end user/stakeholder dalam pemanfaatan hasil keluaran model Anda.

---

Apabila sudah selesai, buat repository baru di github kalian dengan nama Ujian Modul 3 dan *commit* jawaban kalian ke repository tersebut. Akan ada 3 file di repository kalian, yaitu:

- Readme yang proper
- hotel.ipynb
- Dataset

Kirimkan link repository kalian ke **aldo.purwadhika@gmail.com** dan cc ke **operational@purwadhika.com**

Good luck & Happy Coding