

# Laporan Workshop “*Application of Machine Learning in Industry*”

## Big Data Analytics



Disusun Oleh :  
**ANDREAS HARTONO KRISTANTO**  
NRP : 160419047

JURUSAN TEKNIK INFORMATIKA  
FAKULTAS TEKNIK UNIVERSITAS SURABAYA  
SURABAYA

Workshop “*Application of Machine Learning in Industry*” yang dibawakan oleh Bapak Gregorius Edwadr dimana beliau merupakan *machine learning engineer* dimana merupakan bagian dari *software engineer* yang ada di perusahaan *e-commerce* terbesar di dunia yaitu Amazon.com . Dalam workshop ini, bapak Edwadr menjelaskan tentang latar belakang pengalaman bekerja dimana beliau pernah bekerja di GDP labs, Shopee, Agoda, dan terakhir adalah Amazon.com. Beliau pernah bekerja dalam beberapa posisi yang berbeda di beberapa perusahaan seperti posisi *data scientist*, *machine learning engineer*, *data engineer*. Beliau juga menjelaskan perbedaan peran antara *data analyst*, *data engineer*, *data scientist*, dan *machine learning engineer*. Peran *data analyst* adalah memastikan bahwa data yang didapatkan secara mentah atau *raw data* ini diolah sedemikian rupa hingga menjadi sebuah informasi atau laporan kepada pemangku kepentingan dalam sebuah perusahaan.

*Data analyst* tidak memerlukan banyak pengkodean karena lebih mengarah kepada perhitungan statistik. Kemudian peran *data engineer* adalah membuat sebuah infrastruktur untuk proses mentransformasikan data, membersihkan data yang kotor, dan memastikan bahwa *dataset* siap untuk digunakan dalam pengolahan data. Peran *data scientist* lebih bagaimana menemukan sebuah pola atau model baru untuk melakukan *training data*, dan peran *machine learning engineer* adalah seperti dengan *software engineer* pada umumnya yaitu melakukan pengujian kode namun disini lebih spesifik pada *machine learning*.

Dalam contoh penerapan di industri, pak Edwadr juga memberikan contoh dimana ada *client* nya yang meminta dibuatkan sistem rekomendasi seperti aplikasi youtube dimana bisa mengamati perilaku pengguna dari pengguna yang menonton video

tersebut. Disini ternyata hasil akurasi model tidak mengharuskan mencapai angka sempurna yaitu di angka 80%, justru meskipun hasil akurasi model yang diperoleh sekitar 50% namun hal itu sudah cukup untuk membuat client beliau puas dengan hasil yang diperoleh. Disini diperlukan juga bukan hanya hal teknis, namun juga hal - hal lain yang mengarah kepada bisnis, jadi dengan begitu *client* akan lebih mudah untuk mempertimbangkan dalam pengambilan keputusan.

Selain itu, dalam hasil eksperimen saya lakukan pada contoh studi kasus *titanic* dengan tantangan untuk mendapatkan akurasi model mencapai 90% bahkan lebih adalah disini saya mencoba untuk pertama adalah menambahkan jumlah column dan mengubah isi kolom yang ingin digunakan, misal kita meng one-hot dengan tiga column yaitu Pclass, Sex, dan Embarked. Pada kolom yang ingin digunakan, saya mencoba untuk mengubah kolom sex dengan jenis kelamin yang lebih spesifik misalkan wanita, maka saya ubah kolom Sex menjadi Sex\_female, dan kolom Embarked misal menjadi kolom Embarked\_B. Namun hasil yang didapatkan justru tidak memiliki kenaikan akurasi yang signifikan yaitu masih di kisaran 85 % hingga 86 %. Disini saya mencoba untuk mengkombinasikan nya dengan analisis *multivariate* yang dimana merupakan dari eksplorasi analisis data (EDA), namun karena waktu yang terbatas sehingga saya tidak menyelesaikan hasil percobaan dengan kombinasi analisis *multivariate* tersebut.