# PENDAHULUAN

## Latar Belakang Permasalahan

Data telah menjadi asset yang sangat berharga dalam perusahaan, namun yang lebih berhaga dari data adalah informasi yang dapat kita olah dari data tersebut. Informasi yang kita dapatkan dari data memiliki banyak manfaat. Contoh pemanfaatan data menjadi informasi:

Dalam website Forbes dikatakan bahwa salah satu retail mart ingin melakukan strategi penjualan di mana perusahaan tersebut men-target ibu hamil sebagai target pembeli. Ketika ibu hamil maka para ibu membutuhkan keperluan untuk ibu hamil tetapi beberapa waktu kemudian para ibu membutuhkan keperluan untuk bayi yang dilahirkan. Retail mart ini melakukan analisa untuk memprediksi ibu hamil dari data transaksi yang dilakukan pembeli. (Kashmir, 2012)

Walmart, salah satu perusahaan retail terbesar di Amerika telah memiliki banyak cabangan yang tersebar di seluruh Amerika. Perusahaan retail ini harus memenuhi demand atau kebutuhan pelanggan dengan cabang yang buka. Salah satu hal yang mempengaruhi pembelian produk pada Walmart adalah cuaca. Walmart melakukan analisa data penjualan yang digabungkan dengan data cuaca di setiap cabang perusahaan dan menemukan pola penjualan yang dapat membantu mengambil keputusan barang apa yang sebaiknya dijual dengan jumlah banyak sebelum memasuki musin angin topan. (Provost & Fawcett, 2013)

Berdasarkan contoh yang diberikan, data menjadi salah satu asset yang berharga karena jika diolah dengan baik, data dapat menjadi informasi yang sangat baik dalam pengambilan keputusan. Dikutip dari website Wired “When your business is growing, more and more people have opinions about which steps need to be taken. It helps to work with a ‘good data beats opinion’ philosophy. Almost everything can be tested, measured and improved.” (Joris Toonders, 2014), data mengalahkan opini. Data dapat memberikan kenyataan apa yang sedang terjadi, hampir semua pengambilan keputusan yang dilakukan berdasarkan data yang ada, oleh karena itu dapat mengolah data dengan baik dapat memberikan dampak yang baik dalam mengembangkan bisnis terutama dalam bidang Business Intelligence.

Skripsi ini dibuat untuk memberikan sebuah rekomendasi kepada user. Rekomendasi yang diberikan adalah metode analisa yang sesuai dengan karakteristik data yang ingin dianalisa user. Aplikasi akan menerima input berupa file dengan format CSV (*Comma Separated Value*) dan akan memberikan rekomendasi metode analisa yang baik digunakan sesuai dengan input file. Metode analisa yang digunakan adalah Classification dan Regression, kemudian hasil dari proses ini adalah predictive model. Model dapat disimpan kedalam sebuah file.

Pada skripsi ini aplikasi menggunakan beberapa data yaitu West Nile Virus, merupakan data yang berisi mengenai data pengambilan sample nyamuk pada kota Chicago dan indikasi adanya penyakit West Nile Fever pada kota tersebut. House Price, merupakan data yang berisi harga rumah dengan fitur – fitur yang ada pada rumah tersebut. Human Resource, merupakan data yang berisi hasil survey kepuasan pegawai, shift kerja pegawai tersebut, dan cacatan apakah pegawai itu keluar dari perusahaan atau tidak. Insurance, merupakan data yang berisi data pelanggan asuransi mobil beserta harga purchase masing – masing pelanggan. Bakery, merupakan data yang berisi transaksi sebuah toko roti beserta produk yang dijual.

Semua data digunakan untuk analisa metode yang akan direkomendasikan. Lima data ini akan dijadikan input, kemudian user dapat melakukan pre-processing pada aplikasi dan meminta rekomendasi metode pada aplikasi ini. Aplikasi akan memberikan rekomendasi metode dan user dapat memilih untuk mempercayai rekomendasi tersebut atau tidak. Hasil akhir dari aplikasi ini adalah file yang berisi model hasil proses analisa.

## Perumusan Masalah

Masalah yang akan timbul:

1. Bagaimana mengetahui metode yang baik menurut user, contoh: kecepatan atau akurasi.
2. Bagaimana melakukan pemilihan metode – metode *Machine Learning* yang sesuai dengan karakteristik data.
3. Bagaimana melakukan *sampling* untuk melakukan analisa awal untuk menentukan model yang baik

Bagaimana menentukan cara validasi model pada analisa awal yang dilakukan

## Ruang Lingkup

Skripsi ini dibuat dengan batasan:

1. Data yang akan dijadikan input adalah data tabular dengan format CSV (*Comma Separated Value*) yang memiliki kolom lebih dari 2.
2. Banyak  *record* atau jumlah baris pada data tabular lebih dari 1000 baris / *record.*
3. Data CSV sebagian besar diambil dari website http://www.kaggle.com
4. Terdapat 5 Kasus yang akan digunakan untuk analisa, yaitu:
   1. West Nile Virus (https://www.kaggle.com/c/predict-west-nile-virus)
   2. House Price (https://www.kaggle.com/c/house-prices-advanced-regression-techniques)
   3. Human Resource (https://www.kaggle.com/ludobenistant/hr-analytics)
   4. Insurance (https://www.kaggle.com/c/allstate-purchase-prediction-challenge/data)
   5. Bakery (https://wiki.csc.calpoly.edu/datasets/wiki/ExtendedBakery)
5. Sebelum masuk kedalam proses, user dapat memilih menggunakan beberapa kolom saja untuk melakukan analisa
6. Aplikasi dapat menambahkan baris pada data
7. Aplikasi dapat menambahkan kolom pada data dengan nama kolom diinputkan
8. Aplikasi dapat *update* atau merubah data sesuai dengan data yang dipilih oleh user.
9. Aplikasi dapat *delete* atau menghapus data sesuai dengan data yang dipilih oleh user.
10. Aplikasi dapat menggambarkan plot atau grafik
11. Proses pada data menggunakan metode:
    1. *Classification*

Proses klasifikasi data pada suatu kelas seperti (Beli, Tidak Beli)

* 1. *Regression*

Proses menghitung berapa besar nilai yang akan keluar berikutnya seperti berapa besar harga rumah.

1. Proses Rekomendasi, merupakan proses dimana aplikasi akan memberikan rekomendasi metode analisa yang sebaiknya digunakan. Proses ini akan mengambil data *sample* kemudian data tersebut akan dijalankan kepada beberapa metode yang berbeda dan setelah itu melakukan evaluasi untuk menentukan metode yang akan digunakan.
2. User dapat memilih untuk mempercayai hasil rekomendasi atau menolak hasil rekomendasi.
3. Proses *Model Creation* dimana aplikasi akan melakukan analisa dengan motode yang diinginkan user (dari rekomendasi, atau user memilih sendiri)
4. Hasil dari analisa atau output adalah sebuah keterangan atau *result* dari model yang dibuat. Hasil merupakan validasi *accuracy* model.
5. User dapat menyimpan model kedalam bentuk file *pickle (object serialization)*
6. Aplikasi adalah sebuah website

## Tujuan Skripsi

Tujuan dari penelitian ini adalah membuat sebuah aplikasi analisis data dengan fitur rekomendasi metode yang sebaiknya digunakan.

## Metodologi Penelitian

Langkah – langkah dalam pembuatan skripsi:

1. Studi literatur mengenai:
   1. Python – Flask
   2. Cross Validation
2. Desain sistem untuk melakukan proses data beserta rekomendasi metode analisis*.*
3. Analisa Metode untuk menentukan metode yang digunakan dalam sistem rekomendasi
4. Membuat desain sistem untuk menyimpan konfigurasi aplikasi dan konfigurasi user.
5. Membuat tampilan / desain antarmuka.
6. Implementasi Sistem.
7. Evaluasi dan validasi model yang memiliki potensi dengan input data yang ada
8. Pengambilan Kesimpulan
   1. Rekomendasi yang diberikan memiliki *support* yang baik dan membantu pengambilan keputusan*.*
9. Metode yang digunakan tidak menimbukan *overfitting*.

## Sistematika Penulisan

Penulisan laporan Skripsi ini dibagi menjadi beberapa bab, yaitu:

BAB I: PENDAHULUAN

Bab I berisikan judul, latar belakang, perumusan masalah, ruang lingkup, tujuan skripsi, metodologi penelitian, dan sistematika penulisan yang akan digunakan.

BAB II: LANDASAN TEORI

Bab II berisikan teori-teori serta metode-metode yang digunakan dalam pembuatan skripsi.

BAB III: ANALISIS DAN DESAIN SISTEM

Bab III berisikan analisis dan desain sistem yang dibuat

BAB IV: IMPLEMENTASI SISTEM

Bab IV: berisikan tentang implementasi sistem berdasarkan desain sistem seperti pada Bab III.

BAB V: PENGUJIAN SISTEM

Bab V berisikan pengujian sistem yang telah dibuat pada Bab IV.

BAB VI: KESIMPULAN DAN SARAN

Bab VI berisikan kesimpulan yang dapat diambil terhadap hasil yang dicapai, dan saran – saran yang berguna bagi pengembangan selanjutnya.