



PROJECT UAS MACHINE LEARNING

Dosen Pengampu Mata Kuliah: Yefta Christian, S. Kom., M.Kom.

Tim Kuning

1. Andriyanto 1931127
2. Edwin Charley 1931003
3. Steven 1931099
4. Julyan Adi Saputra 1931060

Pendahuluan

Machine Learning merupakan studi tentang algoritma komputer yang meningkat secara otomatis melalui pengalaman dan penggunaan data. *Machine Learning* juga merupakan Ilmu kecerdasan buatan (*Artificial Intelligence*) yang menggunakan teknik statistika untuk menghasilkan suatu model otomatis dari sekumpulan data, dengan tujuan memberikan komputer kemampuan untuk "belajar". Pembelajaran mesin atau *machine learning* memungkinkan komputer mempelajari sejumlah data (*learn from data*) sehingga dapat menghasilkan suatu model untuk melakukan proses *input-output* tanpa menggunakan kode program yang dibuat secara eksplisit. Proses belajar tersebut menggunakan algoritma khusus yang disebut *machine learning algorithms*. Terdapat banyak algoritma *machine learning* dalam efisiensi dan spesifikasi masalah pada suatu usaha salah satunya yaitu UMKM.

Usaha Mikro, Kecil, dan Menengah atau yang biasa disebut sebagai UMKM adalah usaha perdagangan yang dikelola oleh badan usaha perorangan yang merujuk pada usaha ekonomi produktif sesuai dengan kriteria yang ditetapkan oleh UU. Dalam proyek ini, kami bekerja sama dengan mitra UMKM Toko Sifun guna membantu masalah yang dihadapi UMKM saat ini.

Toko Sifun adalah toko swalayan yang bergerak dibidang perdagangan yang menjual barang kebutuhan sehari-hari seperti beras, indomie, gas, air galon, dan lain sebagainya. Toko Sifun berdiri pada tahun 2017 dan terletak di Anggrek Sari blok E1 no 73. Jadi masalah yang dihadapi dalam Toko Sifun yaitu kurangnya pelanggan saat ini dikarenakan sedang dalam pandemi dan kurang luasnya cakupan sistem pemasaran. Dengan ini, kami menerapkan suatu *machine learning* sebagai solusi yang dihadapi UMKM Toko Sifun ini.

Tujuan dari mengimplementasikan *machine learning* terhadap UMKM yaitu kemampuan untuk menyimpan dan mempelajari informasi secara sistematis dapat merangsang perusahaan untuk berkembang, tumbuh, atau meningkat seiring berjalannya waktu sehingga dapat membantu UMKM untuk mengidentifikasi masalah dari pengalaman tanpa perlu intervensi berlebih. Cukup masukan atau input dan ia akan belajar sendiri.

Metode Penelitian

1. Data Preparation

Pertama, kami mengumpulkan data dari Toko UMKM dan membuatnya ke dalam excel dengan format .csv sehingga terbentuk sebuah *dataset*. *Dataset* ini berguna untuk menyelidiki strategi periklanan yang digunakan oleh UMKM kepada kelompok orang mana mereka harus lebih beriklan. Tujuan kami adalah membandingkan pembelian produk oleh orang berdasarkan jenis kelamin, usia, dan perkiraan gaji.

Data ini berisi 5 kolom.

- *UserID* - Setiap orang memiliki ID unik yang darinya kami dapat mengidentifikasi orang tersebut secara unik.
- *Gender* - Seseorang bisa laki-laki atau perempuan.
- *Age* - Usia orang tersebut. Karena produk kami dapat berguna untuk beberapa usia saja.
- *Estimated Salary* - Kolom ini berisi gaji seseorang karena gaji dapat mempengaruhi belanja seseorang.
- *Purchased* - Berisi dua angka '0' atau '1'. '0' berarti tidak dibeli dan '1' berarti dibeli. Variabel ini adalah variabel dependen kami.

Row ID	I User ID	S Gender	I Age	I Estimasi...	S Purcha...
Row0	15624510	Male	19	19000	0
Row1	15810944	Male	35	20000	0
Row2	15668575	Female	26	43000	0
Row3	15603246	Female	27	57000	0
Row4	15804002	Male	19	76000	0
Row5	15728773	Male	27	58000	0
Row6	15598044	Female	27	84000	0
Row7	15694829	Female	32	150000	1
Row8	15600575	Male	25	33000	0
Row9	15727311	Female	35	65000	0
Row10	15570769	Female	26	80000	0
Row11	15606774	Female	26	57000	0

Gambar 1. Data yang digunakan

2. Data Cleaning

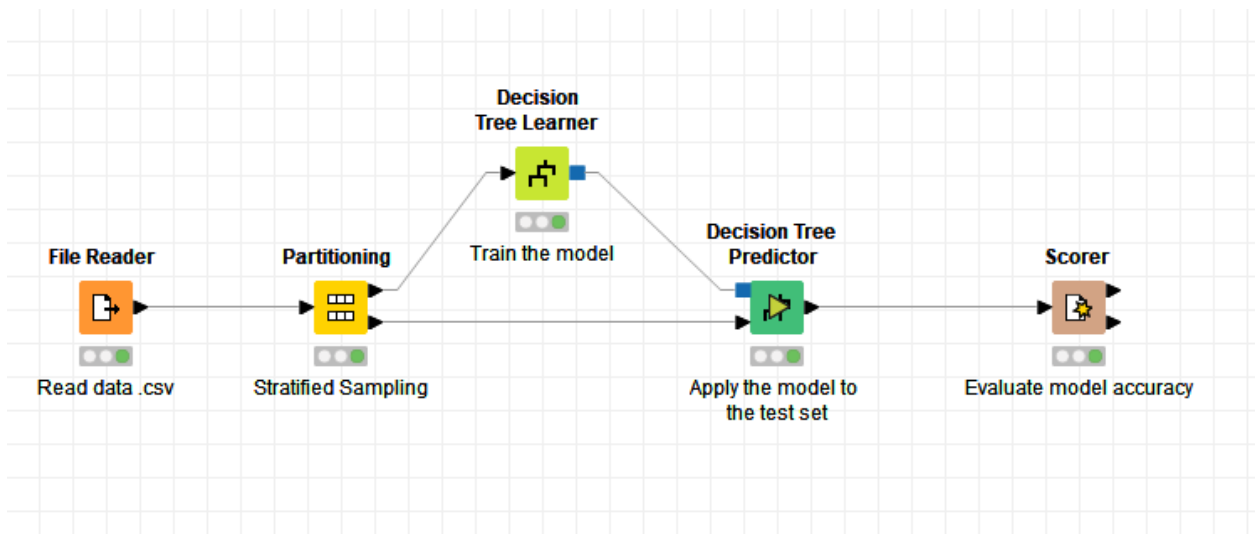
Data Cleaning berarti proses mengidentifikasi bagian data yang salah, tidak lengkap, tidak akurat, tidak relevan, atau hilang dan kemudian memodifikasi, mengganti, atau menghapusnya sesuai kebutuhan. Pembersihan data dianggap sebagai elemen dasar dari ilmu data dasar. Data adalah hal paling berharga untuk *Analytics* dan *Machine learning*. Dalam komputasi atau bisnis data dibutuhkan dimana-mana. Dalam hal data dunia nyata, bukan tidak mungkin data berisi nilai yang tidak lengkap, tidak konsisten, atau hilang. Jika datanya rusak maka dapat menghambat proses atau memberikan hasil yang tidak akurat. Mari kita lihat hasil dari data yang didapat.

i	#False	18
i	#Correct	102
d	Error	0.15
d	Accuracy	0.85

Gambar 2. Pengecekan data

Dari hasil ini bisa kita lihat bahwa akurasi dari evaluasi model akurasi sebesar 85%.

3. Memproses Data dalam Algoritma menggunakan KNIME

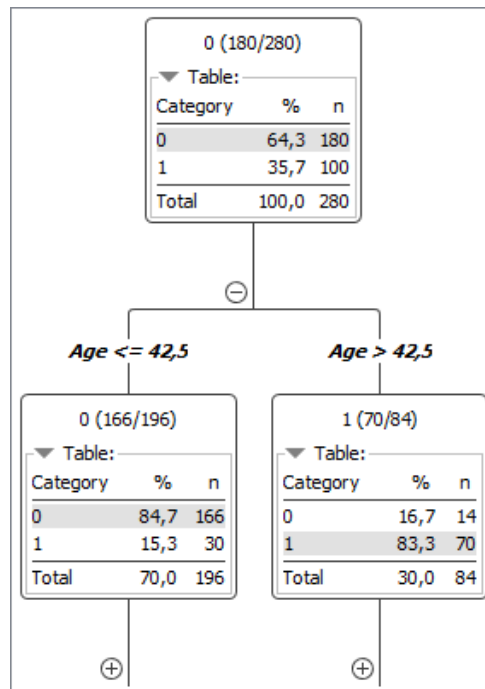


Gambar 3. Proses data menggunakan KNIME

Penulis menggunakan model *tree decision* sebagai model prediksi menggunakan struktur pohon atau struktur berhirarki dikarenakan kemampuannya untuk mem-break down proses pengambilan keputusan yang kompleks menjadi lebih simple, sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan. Dimulai dengan *Tool* pada KNIME yaitu *File Reader*, ini untuk membaca *dataset* salah satunya

dalam bentuk csv. Selanjutnya *Partitioning*, ini untuk melakukan *Stratified Sampling* dimana kami membagi 30% data menjadi set pengujian dan 70% untuk pelatihan. Selanjutnya masuk ke *Decision Tree Learner*, ini untuk melakukan *training* pada model *decision tree*. Selanjutnya masuk ke *Decision Tree Predictor*, disini kita menerapkan model pada set pengujian dan pada *tool* ini hasil dari *tree decision* sudah bisa kita lihat. Selanjutnya *Scorer*, ini berguna untuk mengevaluasi akurasi model. Hasil akurasi yang didapatkan yaitu 85% dari 400 data.

Hasil dan Pembahasan



Gambar 4. Hasil Tree Decision

Berikut hasil dari Algoritma dalam model *tree decision* penulis. Bisa kita lihat dari *tree decision* tersebut, bahwa untuk umur dibawah 42 tahun cenderung tidak melakukan pembelian pada toko, dan yang 42 tahun ke atas lebih banyak melakukan pembelian pada toko tersebut. Jadi untuk sistem pemasaran bisa kita maksimalkan pada target untuk umur ke atas. Dan kita bisa merencanakan strategi pemasaran untuk pelanggan umur 42 kebawah agar penjualan pada toko dapat meluas. Dan berikut hasil dari *F1-Score*, *Recall* dan *Precision* yang didapatkan oleh penulis

Columns: 11	Column Type	Column Index	Color Handler	Size Handler	Shape Han...	Filter Handler	Lower Bound	Upper Bound
TruePositives	Number (integer)	0					31	71
FalsePositives	Number (integer)	1					6	12
TrueNegatives	Number (integer)	2					31	71
FalseNegatives	Number (integer)	3					6	12
Recall	Number (double)	4					0.721	0.922
Precision	Number (double)	5					0.838	0.855
Sensitivity	Number (double)	6					0.721	0.922
Specificity	Number (double)	7					0.721	0.922
F-measure	Number (double)	8					0.775	0.887
Accuracy	Number (double)	9					0.85	0.85
Cohen's kappa	Number (double)	10					0.663	0.663

Gambar 5. Hasil dari F1-Score, Recall dan Precision

Kesimpulan

Machine learning sangat berdampak di segala sektor atau bidang, manfaat pengembangan sistem benar-benar dirasakan oleh masyarakat, terutama generasi milenial. Dalam hal ini *machine learning* memiliki kemampuan untuk memperoleh data yang ada dengan perintah ia sendiri. *Machine Learning* juga dapat mempelajari data yang ada dan data yang ia peroleh sehingga bisa melakukan tugas tertentu. Tugas yang dapat dilakukan oleh *machine learning* pun sangat beragam, tergantung dari apa yang ia pelajari. Penulis menggunakan Machine Learning untuk membantu memecahkan permasalahan pada mitra UMKM. Dan hasil dari *machine learning* bisa digunakan dalam meningkatkan penjualan dan bisa memperluas sistem pemasaran pada UMKM.

Referensi

<https://www.advernesia.com/blog/data-science/machine-learning-adalah/>

<https://towardsdatascience.com/what-is-data-cleaning-how-to-process-data-for-analytics-and-machine-learning-modeling-c2afcf4fbf45>

<https://medium.com/iykra/mengenal-decision-tree-dan-manfaatnya-b98cf3cf6a8d>