

INDIVIDUAL PROJECT

Student Name : Fatimah
Student ID : 1815015049
Course : Data Mining
University : Informatics, Engineering Faculty, UNMUL
Leacture : Mardhiya Hayaty, S.T, M.Kom

Title : Memprediksi Produksi Roti Menggunakan Algoritma KNN

Januari 2022

TABLE OF CONTENTS

TABLE OF CONTENTS.....	ii
BAB I.....	1
1.1. Aims and Objectives	1
1.2. Overview Of The Project	1
BAB II	2
2.1. Dataset	2
2.2. Algorithm	2
BAB III.....	3
3.1. Project Flow	3
3.2. Module.....	3
3.2.1. Pandas	3
3.2.2. Numpy.....	4
3.2.3. Scikit Learn	4
3.3. Result.....	4

BAB I

INTRODUCTION

1.1. Aims and Objectives

Indonesia telah dilanda pandemi sejak Maret 2020 hingga sekarang. Pandemi tidak hanya memberikan dampak besar pada pendidikan melainkan juga pada perekonomian. Hal itu disebabkan masyarakat Indonesia dianjurkan untuk mengurangi kegiatan diluar rumah dan berkerumun. Akibatnya, banyak menimbulkan kerugian terutama pada Usaha Mikro Kecil dan Menengah (UMKM), sebab pengunjung yang datang menurun drastis hingga tidak adanya event atau tempat hiburan yang menjadi sarana mata pencaharian mereka. Selain itu, minimnya penghasilan mereka juga diakibatkan minimnya produksi sebab cuaca yang kurang stabil. Oleh sebab itu, tujuan dari *project* ini adalah memprediksi produksi roti menggunakan Algoritma KNN guna membantu para pengusaha UMKM kedepannya.

1.2. Overview Of The Project

Project yang akan dibuat adalah memprediksi produksi roti dengan kondisi-kondisi tertentu menggunakan algoritma KNN. Hiptotesis awal pada project ini adalah roti akan lebih banyak di produksi jika cuaca sedang baik, di produksi saat weekend serta adanya event. Dataset yang digunakan adalah dataset produksi roti yang terdiri dari 5 kolom dan 6 baris.

BAB II TECHNOLOGICAL BASIS

2.1. Dataset

Dataset yang digunakan adalah dataset dengan format .csv yang terdiri dari 5 *features* (kolom) dan 6 *instance* (baris). Adapun 5 kolom diantaranya adalah kolom kategori roti, kondisi cuaca (*weather*), hari libur atau hari kerja (*holiday*), ada tidaknya event atau tempat hiburan yang menjadi salah satu sarana mata pencaharian (*game*), dan jumlah roti yang di produksi.

Adapun kategori roti yang di produksi sebanyak 6 kategori. Parameter kondisi cuaca yakni 1 – 5, dimana semakin besar nilai parameternya maka semakin baik pula kondisi cuaca. Sedangkan untuk hari libur (*weekend/holiday*) bernilai 1 dan hari kerja (*weekday*) bernilai 0. Jika ada event maka bernilai 0 dan jika tidak ada event maka bernilai 1.

Tabel 2. 1 Dataset

Category	Weather V-1	Holiday V-2	Game V-3	Qty
A	5	1	0	250
B	3	1	1	200
C	1	1	0	75
D	4	1	1	400
E	4	0	0	150
F	2	0	0	50

2.2. Algorithm

Algoritma KNN (*K-Nearest Neighbor*) adalah algoritma *supervised learning* dimana hasil dari *instance* yang baru diklasifikasikan berdasarkan mayoritas dari kategori **k**- tetangga terdekat. Tujuan dari algoritma ini adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan sampel-sampel dari *training data*.

BAB III

METHODS AND APPROACH

3.1. Project Flow

1. Menentukan jumlah tetangga yang akan diperhitungkan (k) yakni $k = 4$
2. Memuat beberapa *library* python yang akan digunakan
3. Memuat dataset yang akan digunakan menggunakan *library* pandas dengan *function* **read_csv**
4. Memasukkan kondisi dengan parameter nilainya dalam variabel bernama *query*, dimana :
 - a. Cuaca = 1
 - b. Weekday = 0
 - c. Game = 0
5. Menghitung *distance* atau kuadrat jarak masing-masing objek terhadap data sampel yang diberikan
6. Mengurutkan objek tersebut dari yang mempunyai jarak lebih dekat
7. Didapatkan 3 kategori terdekat yakni C, F, B, E
8. Menghitung jumlah rata-rata dari 4 kategori terdekat

3.2. Module

3.2.1. Pandas

Pandas adalah library Python yang fokus untuk proses analisis data seperti manipulasi data, persiapan data, dan pembersihan data. Pandas menyediakan struktur data dan fungsi high-level untuk membuat pekerjaan dengan data terstruktur/tabular lebih cepat, mudah, dan ekspresif. Dalam pandas terdapat dua objek yang akan dibahas pada tutorial ini, yaitu *DataFrame* dan *Series*. Pandas memadukan library NumPy yang memiliki kemampuan manipulasi data yang fleksibel dengan database relasional (seperti SQL). Sehingga memudahkan kita untuk

melakukan reshape, slice dan dice, agregasi data, dan mengakses subset dari data.

3.2.2. Numpy

Numpy adalah library salah satu *library* dalam Python yang berfungsi untuk proses komputasi numerik. Hal itu bersesuaian dengan nama NumPy yang memiliki kepanjangan Numerical Python. NumPy memiliki kemampuan membuat objek N-dimensi array. Array merupakan sekumpulan variabel yang memiliki data yang sama. Langkah pertama dalam membuat array adalah dengan cara mengimport terlebih dahulu dengan cara mengetikan `import numpy as np`.

3.2.3. Scikit Learn

Scikit-Learn merupakan library Machine Learning open source berbasis Python yang bisa digunakan dalam Data Science. Kelebihan Scikit-Learn adalah penggunaan API yang mudah serta kecepatannya saat melakukan tolok ukur yang berbeda dalam dataset. Sklearn kompatibel dengan NumPy dan SciPy. Artinya dapat beroperasi dengan library-library yang berbeda untuk Python dengan mudah. Scikit-Learn memberikan sejumlah fitur untuk keperluan Data Science seperti algoritma Regresi, pengelompokan, algoritma Naive Bayes, algoritma Decision Tree, parameter tuning, data preprocessing tool, export/import model, Machine learning pipeline dan algoritma klasifikasi termasuk gradien, K-means, mesin dukungan vektor, DBSCAN, dan juga mampu beroperasi dengan SciPy dan NumPy.

3.3. Result

Berdasarkan output pada prediksi berat badan, didapatkan bahwa jika kondisi cuaca bernilai 1, weekday = 0 dan game = 0 maka roti yang harus dibuat sebanyak 118.75 atau sama dengan 119.