

**PROPOSAL PROYEK AKHIR *DATA MINING***  
**Hubungan Aspek dengan Potensi UMKM di Jawa Timur**



**Disusun oleh:**

Veronika Appeperis	C14200041
Ivan Chandra	C14200119
Felix	C14200165
Princen Kwangtama Tekayadi	C14200181

**PROGRAM STUDI TEKNIK INFORMATIKA**  
**UNIVERSITAS KRISTEN PETRA**  
**2022**

# **BAB 1. INTRODUCTION**

## **1. Latar belakang**

Saat ini, UMKM atau Usaha Mikro Kecil Menengah telah menjadi pilar penopang dari perekonomian bangsa Indonesia. Tercatat UMKM-di Jawa Timur mampu menyerap hingga 97% tenaga kerja se-Jawa Timur dengan kontribusi dalam PDB (Penyumbang Domestik Bruto) mencapai 57,25% jika dilihat di periode 2021/2022. Melihat potensi diatas, ada perlunya untuk menganalisa aspek-aspek yang memiliki pengaruh terhadap potensi UMKM yang dalam hal ini menggunakan klasifikasi UMKM untuk mempermudah pengukuran.

Melalui penelitian ini, akan dilakukan analisis mengenai aspek-aspek UMKM seperti tingkat pendidikan, penggunaan *E-Commerce*, standarisasi, dan izin usaha dengan tingkat pengaruhnya. Dengan ini, dapat diperoleh tambahan wawasan mengenai korelasi aspek-aspek tersebut serta memperoleh rekomendasi-rekomendasi yang dapat mengembangkan potensi pengembangan potensi UMKM di Jawa Timur.

## **2. Rumusan masalah**

- Apakah tingkat pendidikan memiliki pengaruh dengan potensi UMKM di Jawa Timur?
- Apakah penggunaan *E-Commerce* memiliki pengaruh terhadap pengembangan potensi UMKM di Jawa Timur?
- Apakah Standarisasi memiliki pengaruh terhadap potensi UMKM di Jawa Timur?
- Apakah izin usaha berpengaruh terhadap potensi UMKM di Jawa Timur?
- Apakah tingkat pendidikan, penggunaan *E-Commerce*, standarisasi, dan izin usaha memiliki pengaruh signifikan terhadap potensi UMKM di Jawa Timur?

## **3. Tujuan**

- Mengetahui pengaruh tingkat pendidikan pemilik usaha dengan potensi UMKM.
- Mengetahui pengaruh penggunaan *E-Commerce* terhadap potensi UMKM.
- Mengetahui pengaruh Standarisasi terhadap potensi UMKM.
- Mengetahui pengaruh Izin Usaha terhadap potensi UMKM.

#### **4. Manfaat**

Dengan menyelesaikan penelitian ini kami melihat ada beberapa manfaat yang didapatkan, yaitu.

1. Dapat menjadi tambahan wawasan mengenai aspek-aspek yang memiliki pengaruh terhadap potensi UMKM khususnya UMKM di Jawa Timur.
2. Dapat memperoleh rekomendasi bagi pemerintah Jawa Timur agar dapat mengembangkan strategi pengembangan potensi UMKM di Jawa Timur.

## **BAB 2. LANDASAN TEORI**

### **1. Data Preprocessing**

Data preprocessing adalah suatu proses mengubah data mentah menjadi data yang rapi dan efisien. Proses ini diperlukan karena data yang masih mentah biasanya tidak lengkap dan formatnya tidak efisien dan konsisten. Sehingga dengan menggunakan data preprocessing kualitas dari data menjadi baik dan meningkatkan keberhasilan dari penelitian yang melibatkan analisa data . Dengan adanya data preprocessing ini, proyek akan menggunakan proses ini untuk mengubah data mentah UMKM JaTim menjadi data yang efisien dan bersih.

### **2. Correlation Calculation**

Perhitungan Korelasi adalah suatu cara untuk mencari hubungan antara 2 variabel. Dengan analisis korelasi ini, akan digunakan untuk menyatakan tentang derajat antara kekuatan dengan hubungan variabel sehingga dapat mengetahui hubungan variabel dengan mendeskripsikan tentang gambaran yang akan bermanfaat. Melalui proyek ini, akan dilakukan perhitungan melalui atribut - atribut yang memiliki pengaruh terhadap potensi dari UMKM ini.

### **3. Clustering**

Data Clustering adalah suatu proses pengelompokan data kedalam beberapa cluster atau kelompok, proses ini kemudian dilakukan hingga data yang berada dalam satu cluster memiliki tingkat kemiripan yang maksimum. Dengan adanya clustering ini, dapat dilakukan analisa pasar dengan menggunakan pemilihan data dengan atribut - atribut yang memiliki korelasi yang tinggi.

## **BAB 3. METODOLOGI PENELITIAN**

### **1. Desain sistem**

Dalam penelitian ini, dilakukan analisa hubungan antara beberapa attribute seperti tingkat pendidikan pemilik UMKM, penggunaan *e-commerce*, standarisasi, dan izin usaha dengan potensi UMKM yang dalam hal ini menggunakan acuan berupa tingkat klasifikasi dari UMKM. Dengan memanfaatkan dataset yang diperoleh dari lomba “HACKATON 2022 by JCI” berupa “Data UMKM Binaan DISKOPUKMJATIM-1941” dan “Data UMKM Jatim - 975 data” penelitian ini diharapkan dapat memperoleh hubungan dari attribute-attribute di atas.

Metode yang digunakan pada penelitian ini adalah data preprocessing dimana diperlukan untuk melakukan *cleaning* pada dataset dikarenakan data dalam dataset diatas memiliki banyak noise dan ada juga data outlier, proses preprocessing yang dilakukan seperti inputation dan integration . Kemudian dengan data yang telah bersih, dilakukan perhitungan *Correlation Calculation* menggunakan Pearson Correlation untuk menemukan tingkat hubungan antara attribute diatas dengan klasifikasi UMKM. Setelah itu, dilakukan clustering untuk melakukan pengelompokkan pada dataset “Data UMKM Jatim - 975 data”.

Melalui penelitian ini, dapat diperoleh hubungan antara tingkat pendidikan pemilik UMKM, penggunaan *e-commerce*, standarisasi, dan izin usaha dengan potensi UMKM serta tingkat pengaruhnya apakah berpengaruh secara signifikan atau tidak. Dengan ini, selain dapat menambah wawasan mengenai aspek-aspek yang memiliki pengaruh terhadap potensi UMKM khususnya UMKM di Jawa Timur, melalui penelitian ini juga dapat memperoleh rekomendasi bagi pemerintah Jawa Timur agar dapat mengembangkan strategi pengembangan potensi UMKM di Jawa Timur.

### **2. Jadwal kegiatan**

Jadwal kegiatan dari penelitian ini seperti:

1. 18 November 2022 : Mencari Dataset, Define Attribute Meaning, Selecting Dataset
2. 22-29 November 2022 : Membuat Proposal, Data Preprocessing
3. 19 November - 15 Desember 2022 : Data Preprocessing, Correlation Calculation, Clustering, Mencatat hasil penelitian, Visualitation
4. 15 Desember 2022 : Desain Infografis, Presentation Preparation
5. 16 November 2022 : Presentation Preparation, Presentation

### **3. Pembagian Tugas**

- a. Felix  
Mencari Dataset, Selecting Dataset, Membuat Proposal, Data Preprocessing, Correlation Calculation, Clustering, Desain Infografis
- b. Ivan Chandra  
Mencari Dataset, Selecting Dataset, Membuat Proposal, Data Preprocessing, Mencatat hasil penelitian, Desain Infografis
- c. Veronika Appeperis  
Mencari Dataset, Selecting Dataset, Membuat Proposal, Visualization, Presentation Preparation, Presentation, Desain Infografis
- d. Princen Kwangtama Tekayadi  
Mencari Dataset, Selecting Dataset, Define Attribute Meaning, Membuat Proposal, Presentation Preparation, Presentation, Desain Infografis

## BAB 4. HASIL PENELITIAN DAN PEMBAHASAN

### 1. Penjelasan

Dalam penelitian ini, dilakukan analisis hubungan antara beberapa attribute seperti tingkat pendidikan pemilik UMKM, penggunaan *e-commerce*, standarisasi, dan izin usaha dengan potensi UMKM yang dalam hal ini menggunakan acuan berupa tingkat klasifikasi dari UMKM. Dengan memanfaatkan dataset yang diperoleh dari lomba “HACKATON 2022 by JCI” berupa “Data UMKM Binaan DISKOPUKMJATIM-1941” dan “Data UMKM Jatim - 975 data” penelitian ini diharapkan dapat memperoleh hubungan dari attribute-attribute di atas.

Metode yang digunakan pada penelitian ini adalah data preprocessing dimana diperlukan untuk melakukan *cleaning* pada dataset dikarenakan data dalam dataset diatas memiliki banyak noise dan ada juga data outlier, data preprocessing yang dilakukan seperti *cleaning* data dan data *discretization*. Pembersihan data pada dataset “Data UMKM Binaan DISKOPUKMJATIM-1941” yaitu menghapus data yang missing value atau menghapus data-data yang tidak lengkap. Dalam data ini hanya diambil 2 kolom yaitu klasifikasi usaha dan pendidikan terakhir yang ditamatkan. Untuk data *discretization* kami mengelompokkan klasifikasi usaha yang awalnya terdapat ultra mikro, mikro, kecil 1, kecil 2, menengah 1, menengah 2 dan besar menjadi mikro (gabungan dari ultra mikro dan mikro), kecil (gabungan dari kecil 1 dan kecil 2), menengah (gabungan dari menengah 1 dan menengah 2). Data *discretization* juga dilakukan pada kolom pendidikan terakhir yang ditamatkan, awalnya terdapat tidak tamat sd, tamat sd, tamat smp, tamat sma, D3 dan S1 kami kelompokkan menjadi lainnya (gabungan dari tidak tamat sd, tamat sd, tamat smp), tamat sma dan S1 (gabungan dari D3 dan S1).

Pembersihan data yang masih memiliki banyak noise tidak bisa dilakukan dengan sembarang contohnya, di dalam dataset “Data UMKM Jatim - 975 data” walaupun ada data yang *blank* (kosong), data tersebut tidak dihapus dikarenakan dataset “Data UMKM Jatim - 975 data” memiliki banyak data yang *blank* (kosong), jika data yang *blank* (kosong) dihapus akan menyebabkan data dari dataset tersebut hanya tersisa sedikit, sehingga dalam data preprocessing diperlukan keputusan yang terbaik untuk mendapatkan data yang bersih dan tepat. Kemudian dengan data yang telah bersih, dilakukan perhitungan *Correlation Calculation* menggunakan Pearson Correlation untuk menemukan tingkat hubungan antara attribute diatas dengan klasifikasi UMKM. Setelah itu, dilakukan clustering untuk melakukan pengelompokkan pada dataset “Data UMKM Jatim - 975 data”.

Melalui penelitian ini, dapat diperoleh hubungan antara tingkat pendidikan pemilik UMKM, penggunaan *e-commerce*, standarisasi, dan izin usaha dengan

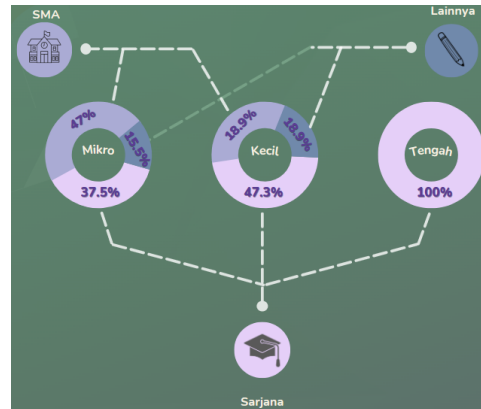
potensi UMKM serta tingkat pengaruhnya apakah berpengaruh secara signifikan atau tidak. Dengan ini, selain dapat menambah wawasan mengenai aspek-aspek yang memiliki pengaruh terhadap potensi UMKM khususnya UMKM di Jawa Timur, melalui penelitian ini juga dapat memperoleh rekomendasi bagi pemerintah Jawa Timur agar dapat mengembangkan strategi pengembangan potensi UMKM di Jawa Timur.



## 2. Hasil Percobaan

Dari penelitian yang dilakukan, dapat diperoleh beberapa hasil dari percobaan dalam mencari aspek potensi UMKM di Jawa Timur, yaitu:

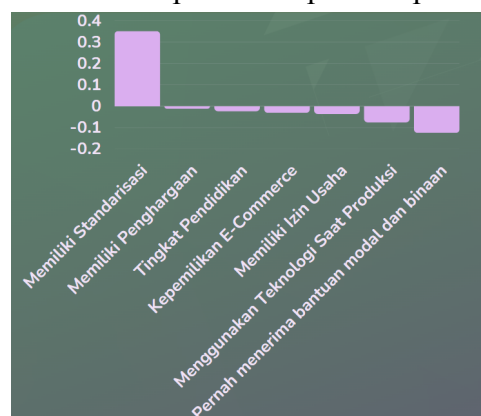
### A. Rasio keterkaitan antara pendidikan dengan klasifikasi usaha



Berdasarkan data UMKM Binaan DISKOPUKMJATIM, ditemukan bahwa terdapat keterkaitan antara tingkat pendidikan pemilik usaha dengan klasifikasi dan potensi dari UMKM, dimana untuk tingkat kecil (47.3%) dan menengah (100%) dipegang oleh Sarjana sedangkan untuk tingkat mikro (47%) dipegang oleh lulusan SMA.

### B. Mendapatkan nilai korelasi dengan klasifikasi usaha.

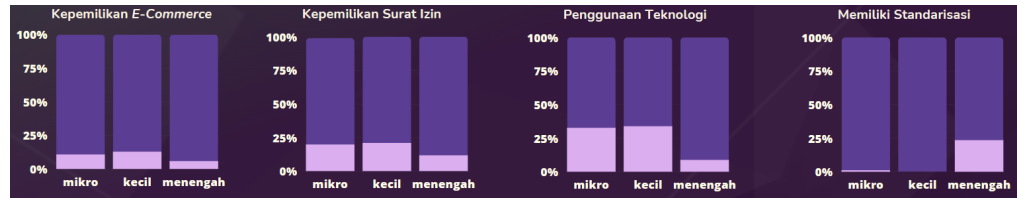
Untuk mendapatkan nilai korelasi dengan klasifikasi usaha, data yang sudah dibersihkan akan difilter terlebih berdasarkan *outlier*. Penelitian ini menggunakan outliers dengan *metric* berupa Euclidean dan jarak neighbors sebanyak 2. Data yang ada di *outliers* akan dipisahkan dengan yang ada di *inliers* untuk mendapatkan data yang lebih rapi dan efisien. Setelah data sudah dipisahkan, data yang ada di *inliers* dimasukan ke dalam function *Correlations* untuk mendapatkan output berupa nilai korelasi dengan klasifikasi usaha.



Dari hasil pengujian klasifikasi tersebut, ditemukan bahwa Klasifikasi usaha memiliki nilai korelasi yang paling tinggi dengan kepemilikan standarisasi sebesar 0.351.

C. Rasio persebaran klasifikasi usaha

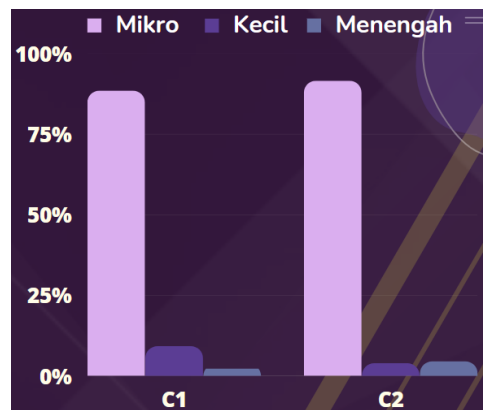
Pada dataset ini, dilakukan pengujian rasio persebaran klasifikasi usaha terhadap 4 aspek, yaitu penggunaan *e-commerce*, kepemilikan surat izin, penggunaan teknologi, dan kepemilikan standarisasi



berdasarkan hasil pengujian tersebut, ditemukan bahwa UMKM kecil memiliki rasio yang paling banyak pada tingkat penggunaan *e-commerce* (12.9%), penggunaan teknologi (33.78%) dan kepemilikan izin (20.97%). Sedangkan UMKM menengah memiliki rasio kepemilikan standarisasi sebanyak 23.53% dibandingkan dengan klasifikasi usaha lain.

D. Mendapatkan hasil clustering berdasarkan aspek klasifikasi usaha

Hasil clustering dari UMKM di Jawa Timur dengan aspek klasifikasi usaha diperoleh dengan menggunakan metode K-Means. Proses pengerjaannya dilakukan dengan memfilter data yang sudah bersih ke dalam *outliers* untuk mendapatkan data *inliers*. Selanjutnya data *inliers* dicluster menggunakan algoritma K-Means yang kemudian membagi data kedalam 2 cluster. Perlu diketahui bahwa pada proses ini digunakan 2 cluster karena jika dibandingkan jumlah cluster yang lain, penggunaan 2 cluster menghasilkan nilai *silhouette* yang terbesar yaitu 0.411 poin.



Hasil clustering ini hanya memiliki perbedaan rasio yang sangat kecil pada tiap clusternya, yang menjadikan clustering tidak sesuai untuk diaplikasikan pada data ini.

## **BAB 5. KESIMPULAN**

Melalui penelitian yang telah dilakukan diatas, dapat diketahui bahwa tingkat pendidikan pemilik UMKM memiliki pengaruh terhadap tingkat klasifikasi dari UMKM, hal ini dibuktikan dari rasio persebaran tingkat pendidikan pada 3 tingkat klasifikasi yang memperlihatkan bahwa pada klasifikasi yang lebih tinggi, rasio tingkat pendidikannya akan lebih didominasi dengan tingkat pendidikan pemilik usaha yang juga tinggi. Selain itu, dalam penelitian ini juga dilakukan perhitungan terhadap 7 aspek (Memiliki Standarisasi, Memiliki Penghargaan, Tingkat Pendidikan, Kepemilikan E-Commerce, Memiliki Izin Usaha, Menggunakan Teknologi Saat Produksi, Pernah Menerima Bantuan Modal dan Binaan) untuk memperoleh tingkat korelasinya dengan tingkat klasifikasi UMKM di Jawa Timur. Perhitungan ini kemudian menghasilkan aspek “Memiliki Standarisasi” sebagai attribute yang memiliki korelasi dengan tingkat klasifikasi UMKM meskipun memiliki nilai korelasi yang masuk kategori rendah yaitu sebesar 0.351 poin. Kemudian dalam penelitian ini juga dilakukan perhitungan rasio persebaran UMKM berdasarkan penggunaan *e-commerce*, kepemilikan surat izin, penggunaan teknologi, dan kepemilikan standarisasi pada setiap tingkat klasifikasi usaha, yang memperlihatkan 23.53% dari UMKM yang memiliki standarisasi adalah UMKM pada tingkat klasifikasi menengah.

Melihat hasil penelitian yang telah dilakukan, dapat diperoleh kesimpulan bahwa untuk meningkatkan potensi UMKM di Jawa Timur, peningkatan kualitas pendidikan di Jawa Timur dapat menjadi salah satu pilihan untuk dilaksanakan karena dengan meningkatnya tingkat pendidikan pemilik UMKM, maka UMKM tersebut akan memiliki potensi tambahan kedepannya. Selain itu, meningkatkan kesadaran pemilik UMKM untuk memiliki standarisasi juga dapat menjadi rekomendasi lain yang dapat dilakukan, karena dengan standarisasi yang baik, pengelolaan dapat lebih terbantu yang kemudian dapat meningkatkan dan mengeluarkan potensi dari UMKM yang dimiliki.

## **BAB 6. REFERENSI**