

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

KTP elektronik (E-KTP) merupakan kartu identitas setiap warga. E-KTP mirip dengan KTP biasa, hanya ditambah *chip* yang berfungsi sebagai *smart card*. E-KTP memuat data Nomor Induk Kependudukan (NIK), nama, tempat dan tanggal lahir, jenis kelamin, agama, status perkawinan, golongan darah, alamat, pekerjaan, kewarganegaraan, foto, masa berlaku, tempat dan tanggal dikeluarkan, tandatangan, serta nama dan nomor induk pegawai pejabat yang menandatanganinya [1]. E-KTP merupakan dokumen elektronik (*e-document*), namun di cetak dalam bentuk kartu. Dokumen elektronik merupakan pelayanan publik yang digunakan untuk menggantikan kertas karena lebih *fleksibel*, pencarian mudah, menghemat ruang, pengarsipan digital, transfer dokumen lebih mudah, keamanan yang lebih baik dan lebih mudah dalam pemulihan data [2].

Ekstraksi informasi adalah suatu proses untuk mengubah informasi tidak terstruktur yang terdapat dalam teks atau citra ke dalam data terstruktur [3]. Pada data citra E-KTP memuat berbagai karakter yang dapat kita ambil dan dijadikan sebagai suatu informasi. Data citra pada E-KTP yang berisi tulisan atau karakter dapat dikenali dengan menggunakan *Optical Character Recognition (OCR)*. OCR merupakan pengenalan pola yang dapat dilakukan untuk menangani masalah pengenalan karakter pada sebuah citra. OCR pada dasarnya merupakan pengenalan sebuah karakter yang digunakan untuk mengidentifikasi citra alphanumeric dari karakter tulisan tangan, *file* ataupun data citra menjadi sebuah teks atau tulisan [4], [5]. Proses OCR dipengaruhi oleh beberapa hal yang dapat menghasilkan akurasi yang baik pada OCR. Secara umum yang dapat mempengaruhi OCR ada dua hal yang utama yaitu mekanisme ekstraksi ciri dan mekanisme pengenalan [4].

Pada penelitian sebelumnya yang terkait dengan ekstraksi informasi pada E-KTP telah dilakukan Sugeng. Penelitian dilakukan pada 50 data E-KTP warga Jawa Timur, berdasarkan pengujian dengan menggunakan algoritma *Template*

*Matching* diperoleh hasil rata-rata dari pengenalan karakter sebesar 90,69% [6]. Penelitian lain yang berkaitan dengan pengenalan karakter yang dilakukan oleh Sandy, Remy, dan Derry pada objek citra Surat Izin Mengemudi (SIM). Berdasarkan hasil pengujian yang dilakukan pada 30 SIM, mendapatkan nilai akurasi sebesar 32,53% untuk perhitungan koefisien korelasi, 48,38% untuk perhitungan *euclidean distance* dan 52,93% untuk perhitungan *manhattan distance*. Penelitian ini menggunakan algoritma *Template Matching* sebagai tahap klasifikasi dan menggunakan *Principal Component Analysis* pada tahap ekstraksi ciri[7]. Peneliti lain yang menggunakan algoritma *Template Matching* mendapatkan hasil pengujian dengan rata-rata 92,90% dan 92,74% [4], [8]. Kemudian penelitian yang dilakukan oleh Angga, Agus, dan Moh Edi untuk melakukan deteksi pada E-KTP dengan menggunakan algoritma *Maximally Stable Extremal Regions* (MSER). Berdasarkan hasil pengujian yang dilakukan memperoleh akurasi 84.49% [9]. Penelitian yang dilakukan oleh Sekar, dan Anastasia untuk mengenali NIK pada E-KTP menggunakan *Invariant Moments of HU dan Intensity of Character* (IoC). Berdasarkan penelitian yang dilakukan mendapatkan akurasi 98,57% pada proses segmentasi [10]. Sedangkan penelitian yang dilakukan oleh Muhammad Fadhlurrohman tentang ekstraksi informasi pada citra E-KTP menggunakan *convolutional neural network*, mendapatkan hasil pengujian akurasi sebesar 66,41% dan akurasi ekstraksi informasi sebesar 83,46% [11].

Menurut Rachman (2012), Huang (2003) dan Byvatov (2003) tingkat akurasi dari *Support Vector Machine* (SVM) lebih baik jika dibandingkan dengan metode regresi logistik, *Artificial Neural Network* (ANN), *Naive Bayes*, *Classification Adaptive Regression Tree* (CART) [12]–[14]. SVM merupakan metode berbasis *machine learning* yang mempunyai performansi tinggi dan dapat diaplikasikan secara luas untuk klasifikasi *hyperplane*. SVM memiliki kekurangan yaitu kurang efisien dalam klasifikasi data dalam jumlah besar. Metode *Reduced Support Vector Machine* (RSVM) diperkenalkan oleh Lee dan Mangasarian (2001) dan dikaji ulang oleh beberapa peneliti, bahwa metode RSVM mempunyai kelebihan untuk menangani klasifikasi data dalam jumlah besar. Metode RSVM

juga mempunyai kelebihan dalam waktu komputasi dan penggunaan memori yang jauh lebih kecil daripada SVM, hal itu dikarenakan RSVM hanya menggunakan sebagian karakteristik dari data yang dipilih untuk proses klasifikasinya [15]–[17]. Penelitian yang dilakukan oleh Suryanto dan Purnami (2015) dengan menggunakan metode RSVM pada tiga tipe data yaitu data untuk kasus linier, lingkaran dan data *checkerboard*. Tipe data tersebut dibagi menjadi 4 kelompok dengan jumlah data berurutan dari 500, 1000, 3000 dan 10000. Hasil yang diperoleh dengan menggunakan metode RSVM mendapatkan tingkat akurasi lebih dari 99% dan waktu yang digunakan untuk proses tersebut relatif singkat. Waktu yang digunakan untuk ketiga tipe data dengan jumlah data yang berurutan yaitu, untuk tipe data linear membutuhkan waktu 0,17 - 16 detik, tipe data lingkaran membutuhkan waktu 0,09 - 15 detik, sedangkan untuk tipe data *checkerboard* membutuhkan waktu 0,20 - 24 detik [18]. Algoritma RSVM pada penelitian tersebut hanya melakukan penelitian pada kasus data yang besar, tidak melakukan penelitian pada kasus yang lebih spesifik.

Berdasarkan permasalahan di atas, penelitian ini menerapkan algoritma RSVM untuk mendapatkan hasil akurasi yang baik pada kasus pengenalan karakter yang terdapat pada data citra E-KTP, karena algoritma RSVM ini belum diketahui tingkat akurasinya jika diterapkan pada kasus E-KTP. Maka dari itu, peneliti mengambil judul “*OPTICAL CHARACTER RECOGNITION MENGGUNAKAN METODE REDUCED SUPPORT VECTOR MACHINE PADA EKSTRAKSI DOKUMEN CITRA KARTU TANDA PENDUDUK ELEKTRONIK (E-KTP)*”.

## **1.2 Identifikasi Masalah**

Penelitian pengenalan karakter pada E-KTP banyak dilakukan dengan menggunakan algoritma *Template Matching* dengan hasil akurasi yang berbeda-beda. Kemudian pada penelitian sebelumnya terdapat permasalahan mengenai tingkat akurasi yang kurang baik pada pengenalan karakter citra E-KTP yang disebabkan karena masih terdapat karakter yang belum di kenali. Hal tersebut dipengaruhi oleh proses *preprocessing* dan ekstraksi ciri yang dilakukan

sebelumnya, sedangkan algoritma RSVM pada penelitian sebelumnya mendapat akurasi yang sangat baik dengan waktu yang relatif singkat pada kasus data besar namun algoritma RSVM belum diketahui tingkat akurasinya jika diterapkan pada kasus pengenalan data citra E-KTP. Oleh karena itu, penelitian ini menggunakan algoritma RSVM untuk proses klasifikasi pada kasus pengenalan karakter yang terdapat pada data citra E-KTP.

Berdasarkan latar belakang dan identifikasi masalah yang telah dijelaskan sebelumnya, maka didapat beberapa point diantaranya sebagai berikut :

1. Terdapat beberapa karakter yang terdapat pada citra E-KTP yang belum dikenali dengan baik.
2. Belum diketahui tingkat akurasi pada kasus pengenalan citra E-KTP dengan menggunakan algoritma RSVM.

### **1.3 Maksud dan Tujuan**

Maksud dari penelitian ini yaitu untuk mengimplementasikan ekstraksi informasi pada E-KTP dengan menggunakan algoritma RSVM. Adapun tujuan dari penelitian ini adalah :

1. Mengenali karakter yang terdapat pada data citra E-KTP dengan menggunakan OCR dengan klasifikasi RSVM.
2. Mengetahui tingkat akurasi yang dihasilkan dari proses klasifikasi dengan menggunakan metode RSVM

### **1.4 Batasan Masalah**

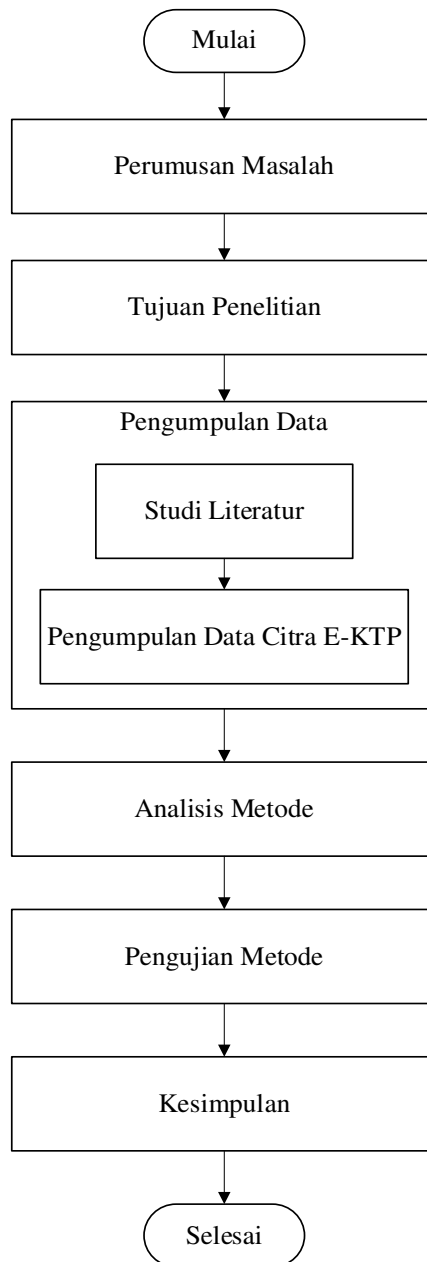
Penelitian mengenai implementasi ekstraksi informasi pada E-KTP ini akan dibatasi agar masalah lebih mengerucut dan mencapai tujuan yang ditentukan, maka batasan masalah dari penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah data citra dari E-KTP sebagai data masukan.
2. Format citra E-KTP adalah *.jpg*, *.jpeg*, dan *.png*.
3. Data yang diambil berupa baris NIK dari citra E-KTP
4. Metode yang digunakan untuk klasifikasi adalah RSVM.

5. Pembangunan sistem untuk pengujian menggunakan bahasa pemrograman *python*.

### **1.5 Metodologi Penelitian**

Metode yang digunakan pada penelitian ini menggunakan metode kuantitatif, di mana metode kuantitatif merupakan metode yang menggunakan hipotesis atau definisi yang telah ditentukan sejak awal penelitian dan dikaji secara kuantitatif. Hasil dari metode kuantitatif menggunakan angka-angka, pengolahan statistik, struktur dan percobaan terkontrol [19]. Metode yang bersangkutan berbanding lurus dengan penelitian yang dikerjakan, di mana penelitian ini untuk mengetahui tingkat akurasi dari hasil klasifikasi pengenalan karakter terhadap E-KTP dengan menggunakan algoritma RSVM. Adapun alur dari penelitian yang dilakukan dapat dilihat pada Gambar 1.1.



**Gambar 1.1 Alur Penelitian**

### **1.5.1 Perumusan Masalah**

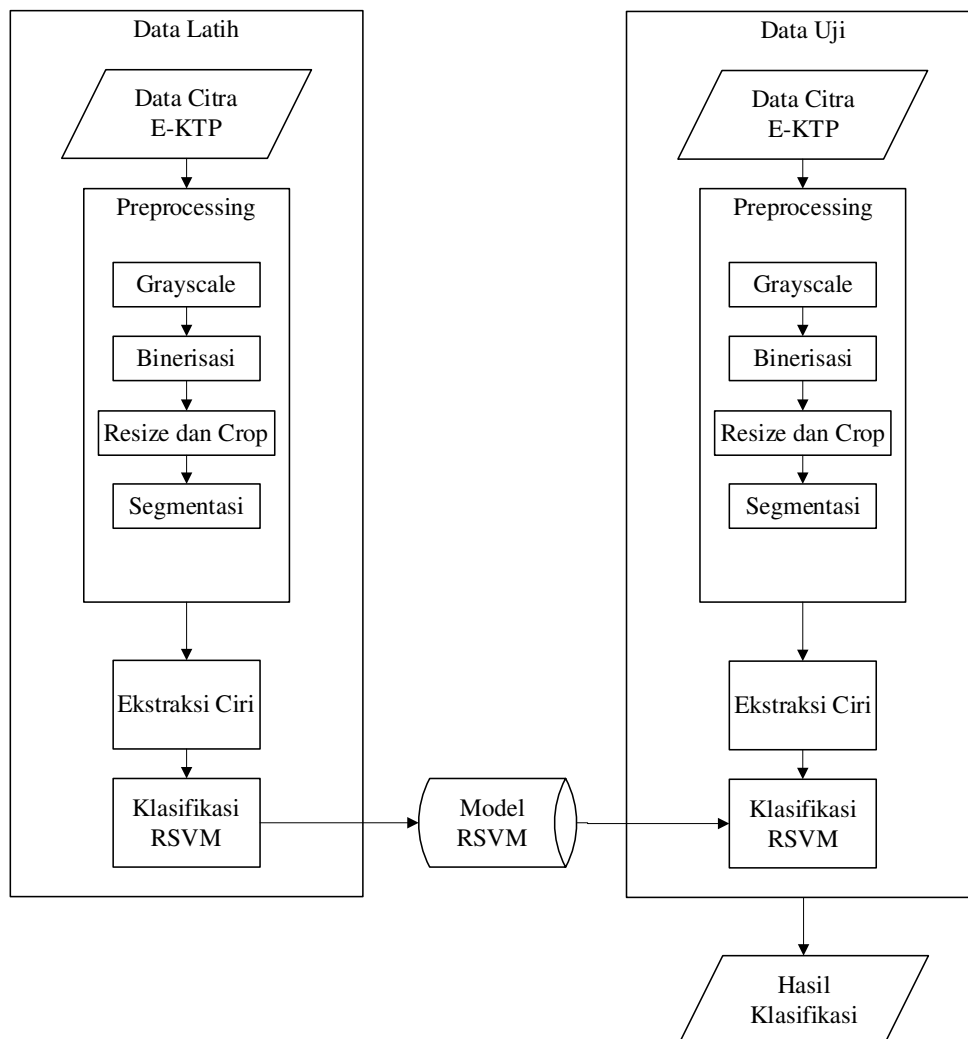
Perumusan masalah ditentukan berdasarkan hasil pengamatan dari penelitian-penelitian sebelumnya mengenai ekstraksi informasi terhadap E-KTP. Sehingga dari hasil pengamatan tersebut dapat dilakukan pengenalan karakter terhadap data citra E-KTP dengan menggunakan metode RSVM.

### 1.5.2 Tujuan Penelitian

Tujuan dari penelitian ini yaitu untuk mengetahui karakter yang terdapat pada citra E-KTP, dan untuk mengetahui tingkat akurasi dari proses klasifikasi dengan menggunakan metode RSVM.

### 1.5.3 Analisis Metode

Analisis metode merupakan penjelasan mengenai metode apa yang akan digunakan pada saat dilakukannya analisis terhadap penelitian. Adapun alur dari tahapan-tahapan analisis metode dapat dilihat pada Gambar 1.2



**Gambar 1.2 Alur Analisis Ekstraksi Citra E-KTP**

Berikut penjelasan tahapan-tahapan dari alur analisis ekstraksi citra E-KTP :

#### 1. Data Citra E-KTP

Data citra E-KTP merupakan data citra masukan yang akan dijadikan sebagai data citra penelitian untuk dilakukan proses pelatihan dan pengujian

#### 2. *Preprocessing*

*Preprocessing* merupakan tahapan yang paling penting, di mana pada tahap ini dilakukan penghapusan pada bagian-bagian yang dirasa tidak diperlukan. Pada *preprocessing*, dilakukan beberapa proses sebagai berikut :

##### 2.1. *Grayscale*

*Grayscale* dilakukan untuk mengubah data citra masukan yang semula berwarna (RGB) menjadi data citra yang mempunyai derajat keabuan yang tinggi atau data citra menjadi abu-abu.

##### 2.2. Binerisasi

Binerisasi adalah proses di mana data citra *grayscale* di ubah menjadi data citra yang mempunyai nilai 1 dan 0 atau bisa juga disebut dengan data citra hitam dan putih.

##### 2.3. *Resize* dan *crop*

Pada tahap ini dilakukan *resize* dan *crop* terhadap hasil dari binerisasi yang kemudian akan dilakukan proses segmentasi. *Resize* dilakukan untuk mengubah ukuran dari data citra agar semua data citra masukan mempunyai ukuran yang sama. Sedangkan *crop* digunakan untuk memotong satu baris NIK.

##### 2.4. Segmentasi

Segmentasi merupakan proses pemisahan area yang diamati pada tiap karakter yang dideteksi.

#### 3. Ekstraksi Ciri

Ekstraksi ciri dilakukan untuk mendapatkan ciri-ciri tertentu dari setiap karakter yang diamati.



#### 4. Klasifikasi RSVM

Pada tahap ini dilakukan untuk mengenali setiap ciri dari karakter yang diamati. Pada proses ini, hasil dari ekstraksi ciri digunakan sebagai pelatihan dengan menggunakan metode RSVM.

### 1.6 Sistematika Penulisan

Sistematika penulisan tugas akhir digunakan untuk mengetahui penjelasan ringkas mengenai penelitian yang dilakukan. Berdasarkan alur sistematika penulisan tugas akhir maka dijelaskan sebagai berikut:

#### **BAB 1 PENDAHULUAN**

Bab ini membahas mengenai Latar Belakang Masalah, Identifikasi Masalah, Maksud dan Tujuan, Batasan Masalah, Metodologi Penelitian, Serta Sistematika Penulisan tugas akhir.

#### **BAB 2 LANDASAN TEORI**

Bab ini membahas berbagai konsep dasar dan teori-teori yang berkaitan dengan Kartu Tanda Penduduk Elektronik, Ekstraksi Informasi, Citra Digital, Pengolahan Citra Digital, *Optical Character Recognition* (OCR), *Preprocessing*, Ekstraksi Fitur, *Reduce Support Vector Machine* (RSVM), dan Bahasa Pemrograman Python.

#### **BAB 3 ANALISIS DAN PERANCANGAN**

Bab ini menjelaskan terkait dengan analisis penelitian seperti Analisis Masalah, Analisis Data, Analisis Metode, dan Analisis Kebutuhan Non Fungsional.

#### **BAB 4 IMPLEMENTASI DAN PENGUJIAN**

Bab ini menjelaskan tahapan Implementasi dan Pengujian dari hasil analisis sebelumnya.

#### **BAB 5 KESIMPULAN DAN SARAN**

Bab ini menjadi bagian akhir dari penelitian yang memuat Kesimpulan dan Saran berdasarkan penelitian yang dilakukan.

