

BAB I

PENDAHULUAN

Pada bab ini menjelaskan bagian pendahuluan penelitian, yang meliputi latar belakang, rumusan masalah, ruang lingkup, tujuan, metode penelitian, tinjauan pustaka, kontribusi penelitian, dan sistematika penulisan.

1.1 Latar Belakang

Pengaruh teknologi digital pada intinya tidak mengubah fungsi *huruf* sebagai perangkat komunikasi visual, namun teknologi komputer menyodorkan beragam spektrum dalam menyampaikan pesan lewat huruf. Seperti contoh, mulai dari desain publikasi (baca: buku, majalah), multimedia, hingga video musik tidak hanya menyampaikan sesuatu pesan atau gagasan yang spesifik, tetapi mencitrakan sebuah gaya yang memiliki korelasi dengan khalayak tertentu, di mana desainer grafis memiliki kebebasan menciptakan visualisasi pesan dengan huruf, tidak hanya untuk dibaca, tetapi juga mengekspresikan suasana atau rasa (Sihombing, 2015). Tipografi itu sendiri, dalam kata lain, lebih merupakan perangkat yang independen, bukan sebagai *PostScript*, bahasa komputer yang digunakan untuk ‘menggambar’ huruf ke dalam kode-kode tipografi.

Dalam dunia tipografi, *Font* adalah sebuah kumpulan karakter lengkap yang mempunyai satu ukuran dan gaya. Sedangkan dalam dunia komputer, *Font* merupakan *file* data elektronik yang mengandung sebuah kumpulan dari elemen penulisan (*glyph*), karakter-karakter atau simbol-simbol. Terdapat lebih dari ratusan ribuan jenis *Font* di dunia, (Agung Haryono, 2014). Sistem operasi ternama di dunia, Microsoft Windows, yang mampu menampung hingga 4000 *Font* dalam sistemnya, menggunakan *TrueType* sebagai format *Font* digital.

TrueType, *PostScript* dan *OpenType* adalah semua standar *font* multi-platform *outline* yang spesifikasi teknisnya tersedia secara terbuka (Phinney, 2001). Perbedaan pertama antara *font TrueType* dan *PostScript* adalah penggunaan berbagai jenis matematika untuk menggambarkan kurva.

Sebagian besar jenis *font* yang digunakan dalam pencet dan di layar milik dua gaya *font* yaitu *font serif* dan *sans-serif* (Altaboli, 2013). *Font* ini dipilih karena mereka adalah dua dari jenis *font* yang banyak digunakan di web dan karena mereka adalah jenis *font* standar dari banyak pengembangan web yang paling populer.

Ketika melihat sebuah desain hasil karya atau aplikasi orang lain, para *graphic designer* dan *frontend developer* terinspirasi untuk membuat hal yang sama, atau menggunakan *font* yang sama. Namun *font* tersebut sudah menjadi gambar atau sebuah aplikasi dan sulit untuk mengetahui apa jenis dari *font* tersebut.

Optical Character Recognition (OCR) merupakan sebuah seni mendeteksi, membagi dan mengenali huruf dari suatu citra. OCR adalah sebuah proses pendeteksian dan pengenalan huruf dari sebuah citra dan mengubahnya menjadi kode *American Standard Code for Information Interchange (ASCII)* atau bentuk lainnya yang dapat dimengerti oleh komputer/mesin (Rao, A.S.C.S.Sastry, A.S.N.Chakravarthy, & P, 2005-2015).

(V & Rani, 2017) melakukan penelitian mengidentifikasi jenis *font* menggunakan metode *Support Vector Machine (SVM)* yang menghasilkan tingkat akurasi rata-rata 80% sedangkan *K-Nearest Neighbor (KNN)* yang menghasilkan tingkat akurasi rata-rata 75,5%.

(Suguna & Thanushkodi, 2010) melakukan penelitian untuk memodifikasi metode *K-Nearest Neighbor* untuk meningkatkan akurasi pada tahap pengklasifikasian dengan cara menambahkan *genetic algorithm* untuk menentukan nilai *k*. Percobaan dan hasil menunjukkan bahwa metode yang digunakan tidak hanya mengurangi kompleksitas *K-Nearest Neighbor*, tetapi juga meningkatkan akurasi klasifikasi.

(Mutrofin, Izzah, Kurniawardhani, & Masrur, 2015) melakukan penelitian yang berjudul *Optimasi Teknik Klasifikasi Modified K Nearest Neighbor Menggunakan Algoritma Genetika* yang bertujuan perbaikan algoritma *Modified K Nearest Neighbor* untuk menentukan nilai *k* optimal. Keunggulan dari algoritma GMKNN

adalah dapat menentukan nilai k optimal pada MKNN dengan otomatis, tanpa harus mencoba satu persatu dalam menentukan nilai k .

Pada penelitian ini dirancang sebuah sistem untuk mengenali jenis *font* dalam bahasa latin, menggunakan metode *Genetic Modified K-Nearest Neighbor (GMKNN)* untuk mengukur akurasi dan waktu. Oleh karena itu, dengan adanya sistem ini diharapkan dapat membantu para *graphic designer* dan *frontend developer* untuk mengetahui jenis *font* pada sebuah desain grafis ataupun sebuah *website*.

1.2 Rumusan Masalah

Hal yang dapat disimpulkan untuk rumus permasalahan penelitian yang diajukan adalah :

1. Bagaimana merancang dan membangun sebuah sistem yang dapat mengidentifikasi jenis *font* yang terdapat pada citra.
2. Bagaimana menerapkan metode *genetic modified k-nearest neighbor* pada sistem identifikasi *font*.
3. Bagaimana mengukur akurasi dan waktu dalam mengidentifikasi *font*.

1.3 Tujuan

Penelitian ini bertujuan untuk mengukur akurasi dan waktu komputasi pada metode *genetic modified k-nearest neighbor* untuk mengidentifikasi jenis *font* pada tahapan klasifikasi citra untuk membantu para *graphic designer* dan *frontend developer*.

1.4 Ruang Lingkup

Penelitian ini berfokus pada mengidentifikasi jenis *font* yang ada pada citra. Adapun ruang lingkup yang dibuat yaitu sebagai berikut :

1. Sistem berbasis *website*.
2. Objek dari citra merupakan teks dengan huruf latin.
3. Huruf yang digunakan pada penelitian adalah jenis *font* dalam keluarga *Sans-Serif*.
4. Satu citra terdapat satu jenis *font*.

5. Citra hasil *training* berukuran 15x15 piksel.
6. *Background* citra polos dan berwarna terang dan *text* berwarna gelap.
7. Gaya *font* yang digunakan adalah *regular*, *italic* dan *bold*.

1.5 Metode Penelitian

Pada tahap awal citra di *input* ke dalam sistem yang terdapat *font* yang ingin diidentifikasi kemudian tahapan selanjutnya citra di proses pada tahapan *pre-processing*, *segmentasi*, *ekstrasi fitur* dan terakhir di *klasifikasikan* yang menghasilkan jenis *font* dari citra yang di *input*. Ketika citra sudah berhasil diambil oleh maka sistem memproses citra. Setelah citra diambil maka pada tahapan selanjutnya adalah tahapan mengubah citra RGB menjadi *grayscale* kemudian menggunakan *low pass filtering* dan *sharpening* untuk memperbaiki kualitas citra setelah memperbaiki kualitas citra tahapan selanjutnya mengubah citra menjadi citra biner menggunakan *otsu thresholding*, kemudian dilakukan proses *mathematical morphology* menemukan lokasi baris, lokasi kata, serta lokasi huruf/karakter dari teks tersebut berada. Setelah menemukan lokasi objek/huruf dilakukan ekstrasi ciri dengan menggunakan algoritma *directional feature extraction* untuk mendapatkan ciri yang dimiliki oleh citra tersebut selanjutnya dilakukan klasifikasi menggunakan metode *genetic modified k-nearest neighbors* dalam tahapan klasifikasi di bagi menjadi dua tahap. Pada tahapan pertama menggunakan algoritma genetika untuk menentukan nilai *k* optimal berdasarkan nilai *validity* kemudian setelah mendapatkan nilai *validity* menentukan tetangga terdekat dengan menggunakan *euclidean distance*. Setelah mendapatkan nilai jarak dilakukan perhitungan *weight voting* pada data test untuk menentukan prediksi kelasnya dilakukan voting berdasarkan besarnya *weight voting* tergantung pada kelas *data training*.

1.6 Tinjauan Pustaka

Dalam penelitian ini melibatkan beberapa tinjauan pustaka yang berkaitan dengan penelitian yang dilakukan adalah :

(Mutrofin, Izzah, Kurniawardhani, & Masrur, 2015) melakukan penelitian yang berjudul *Optimasi Teknik Klasifikasi Modified K Nearest Neighbor Menggunakan Algoritma Genetika* yang bertujuan perbaikan algoritma *Modified K Nearest Neighbor* untuk menentukan nilai k optimal dan kontribusi dalam penelitian ini adalah menggabungkan metode *genetic algorithm* dan *Modified K Nearest Neighbor* yang mendapatkan hasil dengan rata-rata tingkat akurasi 100%.

(V & Rani, 2017) melakukan penelitian yang berjudul *A Font style classification system for English OCR* yang bertujuan untuk mendeteksi jenis font dalam bahas inggris dan kontribusi dalam penelitian ini adalah menambahkan *left distance profile features, right distance profile features* dan *diagonal distance profile features pada pre-proccesing* yang mendapatkan hasil dengan rata-rata tingkat keakurasian 80%.

(Wafiyah, Hidayat, & Perdana, 2017) melakukan penelitian yang berjudul *Implementasi Algoritma Modified K-Nearest Neighbor (MKNN) Untuk Klasifikasi Penyakit Demam* yang bertujuan untuk mengklasifikasikan penyakit demam kontribusi dalam penelitian kali ini adalah menggunakan metode *Modified K-Nearest Neighbor (MKNN)* yang mendapatkan hasil akurasi 88.55% untuk pengujian nilai k, pengujian data latih adalah 92.42%, pengaruh komposisi data latih dan data uji terhadap 96.35%.

(Bhuvaneswari & Therese, 2015) melakukan penelitian yang berjudul *Detection of Cancer in Lung With K-NN Classification Using Genetic Algorithm* yang bertujuan untuk deteksi kanker paru-paru kontribusi dalam penelitian kali ini adalah menggabungkan metode *Genetic Algorithm* dan *K-Nearest Neighbour* yang mendapatkan hasil akurasi 90% dengan pengujian nilai K(50-100).

(Tan, 2018) melakukan penelitian yang berjudul *An Improved KNN Text Classification Algorithm Based on K-medoids and Rough Set* yang bertujuan untuk mengklasifikasikan teks. Kontribusi dalam penelitian ini adalah menyatukan metode *k-nearest neighbors* dan *k-medoids*. Dari penelitian ini dapat disimpulkan KM-RS-KNN memiliki efisiensi klasifikasi dan efek klasifikasi yang lebih baik daripada algoritma KNN tradisional.

(Gupta, Gutierrez-Osuna, Christy, Furuta, & Mandell, 2016) melakukan penelitian yang berjudul *Font Identification in Historical Documents Using Active Learning* yang bertujuan untuk mengidentifikasi jenis *font* pada dokumen sejarah. Kontribusi dalam penelitian ini adalah memodifikasi metode *logistic label propagation*. Dari penelitian ini dapat disimpulkan dengan menggunakan 500 data uji penelitian ini mencapai tingkat klasifikasi 84%.

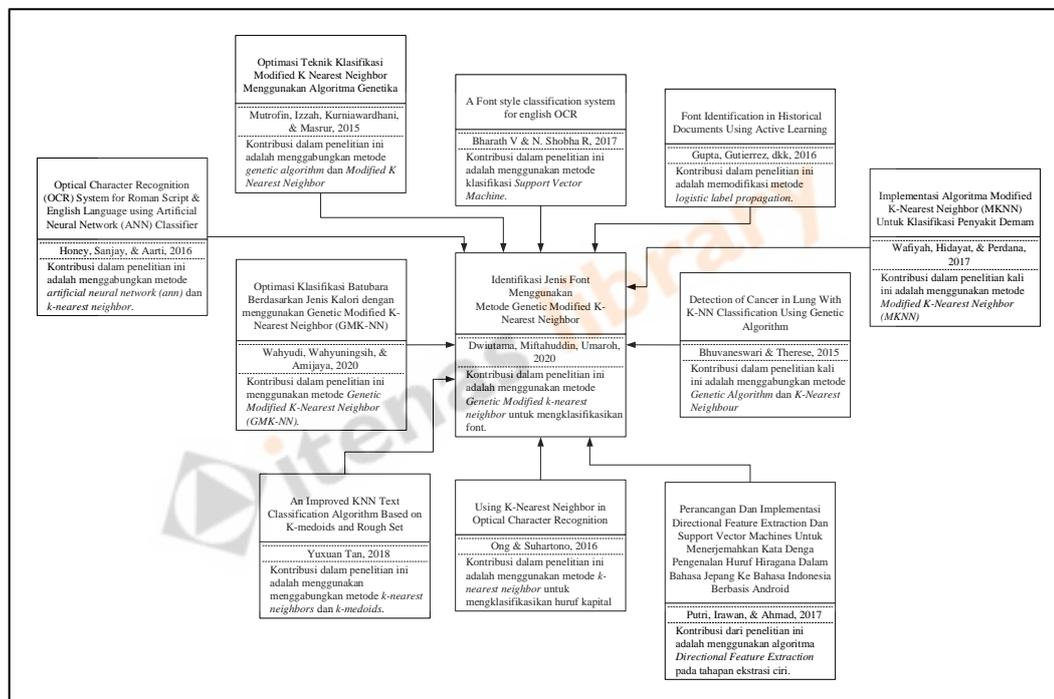
(Honey, Sanjay, & Aarti, 2016) melakukan penelitian yang berjudul *Optical Character Recognition (OCR) System for Roman Script & English Language using Artificial Neural Network (ANN) Classifier* yang bertujuan untuk mendeteksi jenis *font* tertentu. Kontribusi dalam penelitian ini adalah menggabungkan metode *artificial neural network (ann)* dan *k-nearest neighbor*. Kesimpulan dari penelitian ini adalah tingkat keakuratan 98,89% diperoleh dari tiga gaya *font* berbeda.

(Wahyudi, Wahyuningsih, & Amijaya, 2020) melakukan penelitian yang berjudul *Optimasi Klasifikasi Batubara Berdasarkan Jenis Kalori dengan menggunakan Genetic Modified K-Nearest Neighbor (GMK-NN)* yang bertujuan untuk mengklasifikasikan jenis batu bara. Kontribusi dalam penelitian ini menggunakan metode *Genetic Modified K-Nearest Neighbor (GMK-NN)*. Kesimpulan dari penelitian ini adalah akurasi prediksi tertinggi diperoleh dengan menggunakan data proporsi 90:10 yaitu 100%, kemudian dengan proporsi data 80:20 diperoleh akurasi prediksi sebesar 91,67% dan dengan proporsi data 70:30 diperoleh akurasi prediksi 94,44%.

(Putri, Irawan, & Ahmad, 2017) melakukan penelitian yang berjudul *Perancangan Dan Implementasi Directional Feature Extraction Dan Support Vector Machines Untuk Menerjemahkan Kata Dengan Pengenalan Huruf Hiragana Dalam Bahasa Jepang Ke Bahasa Indonesia Berbasis Android* yang bertujuan untuk pengenalan karakter bahasa jepang kemudian menerjemahkannya. Kontribusi dari penelitian ini adalah menggunakan algoritma *Directional Feature Extraction* pada tahapan ekstrasi ciri. Kesimpulan dari penelitian ini adalah menghasilkan nilai akurasi sebesar 85,71%.

(Ong & Suhartono, 2016) melakukan penelitian yang berjudul *Using K-Nearest Neighbor in Optical Character Recognition* yang bertujuan untuk mengklasifikasikan huruf kapital. Kontribusi dalam penelitian ini adalah menggunakan metode *k-nearest neighbor* untuk mengklasifikasikan huruf kapital. Hasil penelitian ini menghasilkan akurasi maksimum 76,9% dengan 200 sampel pelatihan per alfabet.

Penelitian-penelitian tersebut digambarkan keterhubungannya melalui pemetaan pustaka pada Gambar 1.1 berikut.



Gambar 1.1 Peta tinjauan pustaka

1.7 Kontribusi Penelitian

Identifikasi jenis *font* menggunakan metode *genetic modified k-nearest neighbor* dan menambahkan gaya *font* merupakan kontribusi penelitian ini. Penelitian ini mendukung penerapan *smart city* pada bidang *smart living*. Hasil penelitian ini diharapkan dapat memberikan kontribusi bagi para *graphic designer* dan *frontend developer*.

1.8 Sistematika Penulisan

Laporan penelitian ini terdiri dari lima bab dengan sistematika penulisan sebagai berikut :

BAB I PENDAHULUAN

Bab ini menjelaskan mengenai latar belakang, rumusan masalah, tujuan, ruang lingkup, kontribusi penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini menjelaskan analisis berbagai teori dan hasil penelitian yang relevan dengan masalah yang diteliti. Dalam bagian ini melakukan sintesis terhadap teori yang relevan agar diperoleh legitimasi konseptual terhadap variabel yang diteliti.

BAB III METODE PENELITIAN

Bab ini menjelaskan metode yang digunakan dalam penelitian, uraian perancangan dari penelitian yang diusulkan yang meliputi perancangan yang dibutuhkan untuk tahap *input* hingga *output* yang dihasilkan.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Bab ini menjelaskan mengenai implementasi dari perancangan yang telah dilakukan dan pengujian sesuai dengan implementasi yang dikerjakan.

BAB V PENUTUP

Bab ini menjelaskan mengenai kesimpulan dari hasil pengujian pada penelitian ini.