

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

Indonesia merupakan negara dengan keanekaragaman suku bangsa dan budaya. Kebudayaan merupakan keseluruhan sistem gagasan, tindakan, dan hasil karya manusia dalam rangka kehidupan bermasyarakat yang dijadikan milik diri manusia dengan belajar (Koentjaraningrat, 2010). Secara universal, terdapat tujuh unsur kebudayaan meliputi bahasa, sistem pengetahuan, sistem kemasyarakatan atau organisasi sosial, sistem peralatan hidup dan teknologi, sistem mata pencaharian hidup, sistem religi, serta kesenian (Koentjaraningrat, 2010). Kesenian merupakan ekspresi pengalaman keindahan atau pengalaman estetis (Soerjadi-redja, 2016). Di sisi lain, kesenian tradisional itu sendiri merupakan sarana yang digunakan untuk mengekspresikan rasa keindahan dari dalam jiwa manusia dengan latar belakang tradisi atau sistem budaya masyarakat pemilik kesenian (Andri R.M., 2016).

Sebagai salah satu bagian dari kebudayaan, kesenian tradisional Indonesia sangat banyak jumlahnya. Secara keseluruhan, kesenian tradisional yang berkembang di Indonesia terbagi menjadi seni musik, seni tari, seni teater, seni rupa, dan seni sastra (Leonard, 2018). Sejak tahun 2013 hingga 2019, tercatat total 345 kesenian (dalam hal ini seni pertunjukan) sebagai Warisan Budaya Tak Benda Indonesia dan telah diakui, dan jumlah tertinggi terdapat di pulau Jawa dengan total 72 kebudayaan (Kebudayaan, 2020). Seni pertunjukan tersebut berupa seni musik,

seni tari, serta seni pertunjukan itu sendiri (termasuk seni teater di dalamnya). Di luar seni pertunjukan itu, masih banyak kesenian tradisional lainnya yang belum tercatat. Hal ini menunjukkan informasi yang beredar mengenai kesenian tradisional pun akan sangat banyak ragamnya.

Seiring dengan perkembangan jaman, informasi dalam bentuk artikel digital dapat diperoleh dengan mudah melalui teknologi internet, termasuk artikel kesenian tradisional. Dengan begitu banyaknya jenis informasi mengenai kesenian tradisional dan minimnya pengetahuan masyarakat terhadap kesenian tradisional, maka munculnya kesulitan dalam memproses informasi yang ada sangat mungkin ditemukan. Terlebih kebudayaan yang terdapat di pulau Jawa begitu beragam dan memiliki kemiripan satu sama lain. Oleh karena itu, penerapan *Named Entity Recognition* (NER) pada artikel digital kesenian diharapkan dapat membantu pembaca dalam mengidentifikasi istilah-istilah (entitas) yang ada pada artikel, kemudian selanjutnya dapat dicari artinya berdasarkan kamus atau acuan tertentu.

Misalkan terdapat penggalan artikel seperti pada Gambar 1.1, maka NER akan mengidentifikasi istilah-istilah yang ada dan mengelompokkannya ke dalam label yang sudah ada, antara lain dalang, Mahabarata, Ramayana, dan pesinden sebagai tokoh kesenian (mencakup tokoh fiksi, profesi dan nama pelaku kesenian); wayang kulit, wayang, pewayangan sebagai seni pertunjukan; gamelan sebagai alat musik tradisional. Selain tiga label di atas, terdapat dua label lainnya, yaitu musik daerah (terdiri dari judul lagu daerah, misalnya Suwe Ora Jamu, Tokecang, dan Gundhul Gundhul Pacul, serta genre musik daerah seperti Gamelan, Keroncong, dan Gambang Kromong) dan tarian daerah (misalnya Jaipong, Ronggeng, dan Remo).

Wayang kulit dimainkan langsung oleh narator yang disebut dalang. Dalang tidak dapat diperankan oleh sembarang orang. Selain harus lihai memainkan wayang, sang dalang juga harus mengetahui berbagai cerita epos pewayangan seperti Mahabrata dan Ramayana. Dalang dahulu dinilai sebagai profesi yang luhur, karena orang yang menjadi dalang biasanya adalah orang yang terpandang, berilmu, dan berbudi pekerti yang santun.

Sambil memainkan wayang, sang dalang diiringi musik yang bersumber dari alat musik gamelan. Di sela-sela suara gamelan, dilantunkan syair-syair berbahasa Jawa yang dinyanyikan oleh para pesinden yang umumnya adalah perempuan. Sebagai kesenian tradisi yang bernilai magis, sesaji atau sesajen menjadi unsur yang wajib dalam setiap pertunjukan wayang.

Gambar 1.1 Contoh penggalan artikel digital kesenian

(Sumber: [indonesiakaya.com](http://indonesiakaya.com))

Penelitian NER di Indonesia sudah banyak dilakukan, seperti yang dilakukan Suwarningsih dkk (2015) melakukan penerapan NER pada bidang kesehatan. Metode yang diterapkan pada penelitian ini yaitu *Conditional Random Fields* (CRF). Dalam penelitian ini, entitas medis diklasifikasi dalam beberapa jenis, yaitu *location*, *facility*, *diagnosis*, *definition*, dan *person*. Selanjutnya penelitian tersebut diimplementasikan sebagai suatu sistem *medical question answering* dalam bahasa Indonesia. Sebagai data *training*, digunakan 1000 kalimat dari setiap fitur yang ada. Hasilnya, diperoleh nilai akurasi tertinggi 90% dengan menggunakan 3000 kalimat sebagai data *testing*.

Penelitian lainnya berjudul Named Entity Recognition (NER) Bahasa Indonesia Menggunakan Conditional Random Field dan Post-Tagging (Agus Willyawan, 2018). Penelitian tersebut menerapkan metode CRF untuk NER dalam teks bahasa Indonesia, dengan tujuan mempermudah perolehan informasi penting dalam sebuah teks, seperti nama orang, organisasi, dan tempat. Data yang digunakan terbagi menjadi dua yaitu data tidak berlabel yang diperoleh dari

Wikipedia dan data berlabel yang diperoleh dari DBpedia. Dari penelitian ini dihasilkan nilai *precision* dan *recall* di atas 75% dengan nilai F1 sebesar 77%.

Berdasarkan dua penelitian di atas, serta beberapa penelitian rujukan pendukung lainnya, antara lain Named entity recognition model for Indonesian tweet using CRF classifier dengan *precision* 86% untuk *tweet* gabungan (Munarko dkk., 2018); Medical Entity Recognition using Conditional Random Field (CRF) (Herwando, Jiwanggi dan Adriani, 2018); serta Perbandingan Implementasi Conditional Random Fields Dan Random Forest untuk Named Entity Recognition Pada Artikel Berita (Frendy Wong, 2020); maka dilakukan penerapan NER menggunakan metode CRF terhadap kesenian tradisional Indonesia. Adapun data yang diolah yaitu dalam bentuk artikel digital yang diperoleh dari internet. Selain metode CRF, metode lain yang sering digunakan dalam NER yaitu *Hidden Markov Model* (HMM) dan *Maximum Entropy Markov Model* (MEMM), namun CRF dipilih karena merupakan penyempurnaan dari HMM dan MEMM. Selain itu, pada beberapa penelitian yang data *input*-nya juga dalam bahasa Indonesia, CRF mampu meraih nilai akurasi yang lebih tinggi.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang permasalahan yang telah diuraikan, berikut ini merupakan rumusan masalah yang dipilih:

1. Bagaimana cara mengimplementasikan metode *Conditional Random Field* untuk *Named Entity Recognition*?
2. Berapa nilai *f1-score* dari *Conditional Random Field* apabila diterapkan pada artikel digital mengenai kesenian tradisional Indonesia?

### 1.3 Batasan Masalah

Agar pembahasan tidak menyimpang dari latar belakang dan rumusan masalah yang telah diuraikan, maka ditetapkan batasan masalah sebagai berikut:

1. Dataset yang digunakan disusun sendiri, berisi 23.488 kata yang berasal dari 66 artikel berbahasa Indonesia dengan sumber kompas.com. Dataset digunakan untuk *training* dan evaluasi.
2. Data untuk *testing* eksternal berasal dari artikel berbahasa Indonesia dengan sumber Wikipedia.com dan disusun ulang dalam format pdf secara manual.
3. Jenis entitas yang akan diklasifikasi yaitu kesenian yang berasal dari pulau Jawa, berupa tokoh, lagu daerah, alat musik tradisional, tarian daerah, dan kesenian teater sesuai dengan *dataset training*.

### 1.4 Tujuan Penelitian

Tujuan dilakukannya penelitian ini adalah sebagai berikut:

1. Mengetahui cara pengimplementasian *Conditional Random Field* untuk *Named Entity Recognition*.
2. Mengetahui nilai *f1-score* dari *Conditional Random Field* apabila diterapkan pada artikel digital kesenian tradisional di Indonesia.

### 1.5 Manfaat Penelitian

1. Membantu mengklasifikasikan entitas yang ada di artikel digital sesuai dengan jenisnya, sehingga pembaca dapat memperoleh informasi dengan lebih mudah.
2. Sebagai rujukan untuk penelitian selanjutnya

## **1.6 Sistematika Penulisan**

Sistematika penulisan yang diterapkan dalam penyusunan skripsi ini terdiri dari 5 bab utama sebagai berikut:

### **BAB 1 PENDAHULUAN**

Bab pertama menjelaskan latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan laporan.

### **BAB 2 LANDASAN TEORI**

Bab kedua menjabarkan teori-teori yang mendasari penelitian ini, antara lain Named Entity Recognition, Preprocessing, dan Conditional Random Fields (CRF).

### **BAB 3 METODOLOGI PENELITIAN**

Bab ketiga berisi tahapan metode penelitian yang dilakukan disertai dengan diagram dalam bentuk *flowchart*.

### **BAB 4 HASIL DAN DISKUSI**

Bab keempat membahas implementasi kode, hasil pengujian, dan evaluasi terhadap hasil yang diperoleh.

### **BAB 5 SIMPULAN DAN SARAN**

Bab kelima berisi kesimpulan dari penelitian yang telah dilakukan dan saran untuk pengembangan penelitian lebih lanjut