

BAB I

PENDAHULUAN

1.1 Latar Belakang

Machine Learning (ML) didefinisikan sebagai mesin pembelajar dari cabang Artificial Intelligence yang sasarannya adalah mempelajari data sehingga dapat memfokuskan pengembangan sistem. ML dapat belajar secara “independen” tanpa memerlukan manusia untuk melakukan pemrograman berulang kali[1]. *Machine learning* merupakan kode pemrograman yang membaca data lalu mempelajari data. Data yang diproses oleh *machine learning* berjumlah ribuan hingga ratusan ribu. Dalam mengelola data tersebut *machine learning* membutuhkan metode yang efektif untuk mempelajari data, metode yang digunakan bisa menggunakan metode *machine learning* dan *deep learning*. Adapun pengertian dari *Deep learning* sendiri adalah komponen dari pembelajaran mesin (*machine learning*) yang berhubungan dengan algoritma. Algoritma memiliki cara kerja yang disebut dengan jaringan saraf tiruan karena prinsipnya meniru struktur dan fungsi otak [2]. Pembelajaran mendalam (*Deep learning*) meniru struktur otak manusia dengan menggunakan neuron yang memiliki beberapa *layer* sesuai kebutuhan. Jaringan saraf tiruan (*Neural Network*) merupakan dasar dibangunnya model *deep learning*. Riset mengenai *Neural Network* sudah ada sejak tahun 1980 namun dikembangkan lagi saat ini ketika komputer sudah semakin cepat. Selain itu, dengan keberadaan teknologi pada *Big Data* seperti spark berbasis *multi node cluster* dan hadoop, serta pemrosesan secara paralel berbasis GPU, dan sebagainya juga berperan besar dalam pengembangan *Neural Network*[1]. Sebelum masuk pada metode *machine learning* dan *deep learning* diperlukan data untuk melatih metode-metode yang ada pada *machine learning* dan *deep learning*. Salah satu cara untuk mendapatkan data dengan melakukan *mining* atau *crawling* data pada suatu platform.

Data Mining didefinisikan sebagai penguraian atau eksplorasi data dari informasi yang masif serta masih belum teridentifikasi. Namun, hal tersebut dapat dimengerti dan memiliki manfaat dari *database* yang masif serta dimanfaatkan sebagai pembuat keputusan bisnis yang penting [3]. Seperti yang dijelaskan *data mining* adalah penguraian atau eksplorasi data, untuk mendapatkan data yang masif biasanya di dapat dari media sosial dan platform tertentu sesuai dengan kebutuhan data yang diperlukan. Data yang didapatkan setelah itu akan dilakukan *preprocessing* dan pemrosesan sesuai dengan topik yang diperlukan, seperti *clustering*, klasifikasi, regresi dan lainnya. Salah satu yang bisa dilakukan

saat mendapatkan data dari social media atau dari platform lainnya adalah melakukan klasifikasi artikel berita.

Klasifikasi mampu mengelompokkan masing-masing detail suatu *dataset* ke dalam kelas-kelas atau kelompok yang telah didefinisikan sebelumnya [4]. Tujuan dari klasifikasi adalah memprediksi kelompok kata pada suatu data. Sebelumnya pada tahun 2020 Arliyanti Nurdin dkk melakukan penelitian klasifikasi teks menggunakan *word embedding word2vec, glove, fasttext* dan CNN. Word2vec dan GloVe kurang efisien dalam merepresentasikan vektor dari suatu kata yang tidak berada di dalam korpus (out of vocabulary) sedangkan FastText bisa mengatasi permasalahan *out of vocabulary* ini[5]. Dalam melakukan klasifikasi banyak algoritma yang dapat digunakan, contohnya adalah *artificial neural network* (ANN) yang menirukan cara kerja jaringan saraf manusia.

Berdasarkan hasil dari penelitian sebelumnya [6] mengenai penggunaan metode ANN untuk mengklasifikasi opini publik terhadap covid-19. Dari penelitian tersebut, didapatkan hasil perhitungan *accuracy* sebesar 88,62%, *precision* 91,5% dan *recall* 95,73% dengan persentase 80:20 (80% data latih dan 20% data uji) dengan menggunakan 1 *hidden layer* yang berjumlah 10 *node* dan *learning rate* sebesar 0.02. Pada penelitian mengenai analisis sentiment di twitter terhadap kenaikan uang kuliah tunggal menggunakan *Artificial Neural Network* (ANN) oleh Achmad Chikham Nouriel Rosyadi, dari seluruh uji coba yang dilakukan hasilnya memiliki kinerja yang cukup baik, 84% ialah nilai rata-rata hasil dari akurasi presisi, *recall*, dan *f1-score* [7]. Dalam melakukan klasifikasi teks menggunakan *Artificial Neural Network* (ANN) dibutuhkan *layer*, adapun *layer* yang digunakan adalah *input layer, hidden layer* dan *output layer*. Terdapat *neuron* di dalam *layer* yang memiliki nilai salah satunya pada *hidden layer* dan *output layer*, untuk menentukan nilai terhadap *neuron* di *layer* tersebut dibutuhkan yang namanya fungsi aktivasi.

Fungsi aktivasi eksklusif digunakan dalam jaringan saraf tiruan untuk mempelajari pola rumit dalam data [8]. Fungsi aktivasi digunakan untuk menghitung nilai *neuron* pada *layer* berikutnya dimana nilai *neuron* sebelumnya sudah ditentukan dan nilai bobot sudah ada. Beberapa fungsi aktivasi yang bisa digunakan seperti fungsi aktivasi *linear, step, sigmoid, relu, softmax* dan *gaussian*. Manfaat dari fungsi aktivasi adalah mengurangi nilai *error* pada *output* yang dihasilkan pada masing-masing *neuron* [9].

Berdasarkan hasil penelitian yang sudah dilakukan, *Artificial Neural Network* (ANN) memiliki efektifitas dalam melakukan klasifikasi dan memiliki akurasi yang cukup bagus. Penggunaan *Artificial Neural Network* (ANN) dalam topik ini memiliki potensi untuk mengelola data teks pada artikel.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini yaitu sebagai berikut:

1.2.1 Permasalahan

Berdasarkan latar belakang diatas adanya metode klasifikasi teks menggunakan *Artificial Neural Network* (ANN) dapat menjadi solusi untuk melakukan klasifikasi teks pada data yang digunakan. Karena data di kategorikan dengan cara manual oleh dinas wisata dan budaya bangkalan dengan kategori seperti, wisata alam, wisata budaya, dan wisata buatan. dengan menggunakan metode *Artificial Neural Network* (ANN) berpotensi bisa memaksimalkan klasifikasi teks pada berita pariwisata dengan mempelajari pola data yang kompleks.