Perkembangan teknologi di era globalisasi saat ini sangatlah pesat. Perkembangan teknologi ini menyebabkan banyaknya data-data yang tersebar di dunia maya. Penyebaran data-data tersebut bisa melalui melalui sms, media sosial, maupun melalui *website- website* yang ada di dunia internet.

Data-data yang tersebar dalam dunia maya bukan hanya sekitar 100 data atau 1000 data, melainkan jutaan atau mungkin bisa mencapai milyaran atau triliunan data. Data-data tersebut belum terklasifikasikan sesuai dengan topik apa yang dibahas dalam data tersebut. Hal ini akan membutuhkan waktu yang lama jika ingin melakukan pencarian sebuah data yang berisi mengenai satu topik tertentu. Maka dari itu diperlukanlah pengklasifikasian data sesuai dengan topik-topik yang dibicarakan oleh data-data tersebut.

Proses penglasifikasian dapat dilakukan dengan berbagai macam metode pengklasifikasian yang sebelumnya sudah dilakukan oleh para peneliti. Metode-metode itu seperti metode *k-means,* *naïve bayes classifier, k-nearest neighbor, support vector machine (SVM), artificial neural network, fuzzy k-nearest neighbor,* dan lain sebagainya. Dari semua metode-metode yang ada, dipilihlah metode *naïve bayes classifier* dan *k-nearest neighbor* untuk membantu penelitian dalam mengklasifikasikan data-data tersebut. Pemilihan kedua algoritma ini dikarenakan algoritma ini memiliki nilai akurasi yang tinggi. Berdasarkan penelitian yang dilakukan oleh Hamzah didapat bahwa akurasi algoritma *naïve bayes classifier* sebesar 91% pada klasifikasi dokumen berita dan berdasarkan penelitian yang dilakukan oleh Samuel dkk didapatkan akurasi algoritma *k-nearest neighbor* sebesar 88,29% pada klasifikasi subtopik berita, selain itu kedua algoritma tersebut mudah untuk dipelajari.

*K-Nearest Neighbour* atau biasa disingkat dengan istilah KNN biasa digunakan dalam proses klasifikasi, salah satunya dalam klasifikasi teks. *K-Nearest Neighbour* merupakan algoritma yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data yang lain [1]. Untuk mengetahui jarak antar data baru dengan data lama dapat dilakukan dengan menghitung bobot data-data tersebut berdasarkan fitur yang ada, kemudian bobot tersebut akan dicocokan dan dihitung jaraknya.

*Naïve Bayes* merupakan suatu metode yang digunakan untuk mengklasifikasikan suatu objek berdasarkan perhitungan probabilitas. *Naïve Bayes* akan memprediksi suatu objek dengan memperhitungkan probabilitas suatu bukti untuk dijadikan sebagai hipotesis.

Penelitian kali ini akan membuat sistem yang dapat mengambil konten *website* berita *online* secara otomatis yang kemundian akan mengklasifikasikannya ke dalam topik tertentu. Untuk proses klasifikasi itu sendiri akan menggabungkan kedua algoritma diatas, yaitu *K-Nearest Neighbour* dan *Naïve Bayes*. Penelitian kali ini mengambil judul **“Implementasi *Naïve Bayes* dan *K-Nearest Neighbour* dalam Kategorisasi Artikel Berita di Internet”.** Pengambilan berita dari beberapa situs berita ini bertujuan untuk mengklasifikasikan berita-berita yang berasal dari beberapa situs berita sesuai dengan topik yang dibicarakan dalam artikel berita tersebut.