

1. Pendahuluan

Latar Belakang

Pada saat ini, forum-forum *online* sudah banyak bermunculan, terutama yang membahas tentang ulasan sebuah produk dari *user*. Dari ulasan produk yang ditulis user dapat diambil manfaatnya, seperti untuk *market intelligence management*, *enterprise content management*, dan sebagainya [1]. Salah satu elemen penting yang dapat diambil dari teks ulasan adalah pengenalan nama produk. Tetapi, karena teks ulasan biasanya berbentuk teks informal, maka variasi penyebutan untuk nama sebuah produk pun bermacam-macam, sehingga untuk mengekstrak nama produk dengan pencocokan yang benar-benar pasti/*exact match* ke daftar nama merek itu sangat sulit. Salah satu contohnya adalah penulisan Nutrimax Glutoxidant yang merupakan nama sebuah produk kecantikan, di mana produk tersebut bisa ditulis dengan Nutrimax Glutoxidant, Nutrimax Glutox, atau Glutox saja.

Oleh karena itu, dilakukan riset *product named entity recognition* (PRO NER). PRO NER adalah pengenalan entitas bernama yang berfokus pada nama dari sebuah produk. Entitas bernama produk memiliki perbedaan dengan entitas bernama biasa [2]. Ciri-cirinya adalah sebagai berikut. Pertama, entitas bernama produk biasanya terdiri dari merek produk dan tipe produk, di mana keduanya sangat penting dan menjadi pembeda dengan entitas bernama biasa. Kedua, entitas bernama produk memiliki beberapa ekspresi lain yang menjelaskan atribut dan kategori dari produk, yang berperan sebagai elemen tambahan. Contoh penerapannya adalah pada nama produk “MAC 219 Brush”, di mana kata “MAC” adalah merek produk, “219” adalah tipe produk, dan “Brush” adalah jenis produk.

Dalam melakukan PRO NER, terdapat beberapa metode, yaitu *Conditional Random Fields* (CRF) [3], *parsing-based decision list* [4], dan *Hidden Markov Model* (HMM) dan algoritma Viterbi [1]. Pada tugas akhir ini akan menggunakan metode HMM dan algoritma Viterbi. HMM adalah model yang diasumsikan sebagai proses Markov dengan *hidden state*. Sedangkan algoritma Viterbi adalah algoritma yang berfungsi untuk mencari *hidden state* pada HMM [5].

Sebelumnya sudah ada beberapa pekerjaan PRO NER yang telah dilakukan. Niu (2003) telah meneliti NER pada teks bahasa Inggris menggunakan pendekatan *bootstrapping* dan menghasilkan *F1-score* sebesar 69,8% untuk entitas bernama produk [4]. Setelah itu, Liu (2005) yang meneliti PRO NER untuk teks bahasa Cina dengan menggunakan HMM dan algoritma Viterbi, menghasilkan *F1-score* sebesar 79.7%, 86.9%, dan 75.8% untuk masing-masing label produk (PRO), tipe (TYP), dan merek (BRA) [1]. Karena berdasarkan berbagai penelitian tersebut, penelitian Liu yang menggunakan metode HMM-Viterbi memiliki skor yang tertinggi dan melebihi 75% untuk ketiga label PRO NER, maka metode tersebut dapat dianggap cukup baik dan akan diimplementasikan untuk teks berbahasa Indonesia.

Topik dan Batasannya

Penelitian tugas akhir ini mencoba membangun sebuah model pengenalan nama produk pada *posting* di forum *online* menggunakan HMM dan algoritma Viterbi dan mengukur performansi dari model tersebut. Forum *online* yang digunakan adalah forum.femaledaily.com, yaitu forum *online* berbahasa Indonesia yang membahas tentang kecantikan. Dari beberapa *posting* di forum *online* tersebut, telah dibuat dataset yang sudah dilengkapi dengan label PRO NER di setiap katanya [6]. Dalam penelitian ini, dataset tersebut digunakan dan dibagi menjadi data latih dan data uji, dan data latih diproses dengan model yang dibuat sehingga dapat menghasilkan label PRO NER prediksi pada data uji. Label-label yang dapat diprediksi adalah produk (PRO), merek (BRA), dan tipe (TYP).

Tujuan

Tujuan penelitian tugas akhir ini adalah untuk menghasilkan sebuah model prediksi PRO NER yang dibuat dengan HMM dan algoritma Viterbi serta dapat menunjukkan performansi dari model tersebut dalam melabeli data.

Organisasi Tulisan

Penulisan tugas akhir ini mempunyai urutan sebagai berikut. Bagian 2 adalah penelitian-penelitian PRO NER sebelumnya yang berkaitan dengan penelitian ini. Sedangkan bagian 3 adalah gambaran umum, urutan, dan penjelasan dari sistem PRO NER yang dibangun. Pada bagian 4, dijelaskan eksperimen yang dilakukan dan hasil yang didapat beserta analisisnya. Dan pada bagian 5, dijelaskan kesimpulan dari penelitian tugas akhir ini beserta saran yang dapat digunakan untuk penelitian selanjutnya.