

INTISARI

Penelitian ini membahas adaptasi *Automatic Intent Detector (AID)* yang dikembangkan oleh Ferrera *et al.* (2025) pada data komentar *live shopping e-commerce* berbahasa Indonesia yang bersifat *informal* dan tidak terstruktur. Tujuan penelitian ini adalah mengevaluasi kinerja *pipeline AID* dalam konteks percakapan nyata serta mengidentifikasi keterbatasan pelabelan *intent* otomatis pada domain *e-commerce*. Variabel yang dikaji meliputi kondisi *preprocessing* linguistik, representasi *embedding*, kualitas *clustering*, stabilitas *parsing*, dan metode pelabelan otomatis. Metode analisis yang digunakan mencakup representasi *embedding* berbasis *transformer*, analisis kecenderungan *cluster*, *clustering* berbasis HDBSCAN, serta evaluasi kuantitatif dan kualitatif terhadap hasil *clustering* dan pelabelan otomatis.

Hasil penelitian menunjukkan bahwa penerapan *preprocessing* linguistik berupa normalisasi kata *informal* dan *entity masking* meningkatkan kualitas *clustering* yang ditunjukkan oleh peningkatan *Silhouette Score* serta meningkatkan kestabilan *parser* dalam mengekstraksi pasangan *action-object*. Evaluasi pelabelan otomatis menunjukkan bahwa pendekatan *majority action-object* rentan terhadap bias frekuensi, sedangkan pendekatan *centroid-based representative selection* mampu menghasilkan label yang lebih spesifik namun masih memiliki keterbatasan dari sisi keterbacaan.

Kesimpulan penelitian ini menunjukkan bahwa *pipeline AID* dapat diadaptasi secara efektif untuk data komentar *e-commerce* berbahasa Indonesia pada tahap *clustering*. Namun demikian, pelabelan *intent* otomatis masih menjadi tantangan utama, sehingga pengembangan strategi pelabelan yang lebih adaptif diperlukan untuk meningkatkan interpretabilitas *intent* pada konteks percakapan nyata.

Kata kunci: *Automatic Labeling, E-commerce Comments, Entity Masking, Normalisasi Informal, Unsupervised Intent Detection*

ABSTRACT

This study examines the adaptation of the Automatic Intent Detector (AID) developed by Ferrera et al. (2025) to Indonesian-language live shopping e-commerce comments data that is informal and unstructured. The purpose of this study is to evaluate the performance of the AID pipeline in the context of real conversations and to identify the limitations of automatic intent labeling in the e-commerce domain. The variables examined include linguistic preprocessing conditions, embedding representation, clustering quality, parsing stability, and automatic labeling methods. The analysis methods used include transformer-based embedding representation, cluster tendency analysis, HDBSCAN clustering, and quantitative and qualitative evaluation of clustering and automatic labeling results.

The results show that the implementation of linguistic preprocessing consisting of informal word normalization and entity masking improves clustering quality, as indicated by an increase in the Silhouette Score, and increases the stability of the parser in extracting action–object pairs. The automatic labeling evaluation shows that the majority action–object approach is prone to frequency bias, while the centroid-based representative selection approach is capable of producing more specific labels but still has limitations in terms of readability.

The conclusion of this study shows that the AID pipeline can be effectively adapted to Indonesian-language e-commerce comments data at the clustering stage. However, automatic intent labeling remains a major challenge, so the development of more adaptive labeling strategies is needed to improve the interpretability of intent in real conversation contexts.

Keywords: Automatic Labeling, E-commerce Comments, Entity Masking, Informal Normalization, Unsupervised Intent Detection