

DETEKSI SITUS JUDI ONLINE SECARA REALTIME BERBASIS TEXT PROCESSING MENGGUNAKAN ALGORITMA LSTM

Zaidan Jibrán Azhar, 202231116

Dibimbing oleh Dr. Ir. Abdul Haris, S.Kom., M.Kom., IPM.

ABSTRAK

Perputaran judi online di Indonesia mencapai ratusan triliun rupiah per tahun, sementara pemblokiran berbasis daftar alamat cepat usang karena situs mudah berganti domain. Penelitian ini bertujuan merancang dan menguji sistem deteksi situs judi online berbahasa Indonesia yang bekerja secara real-time di peramban dengan menilai isi teks halaman, bukan alamatnya. Cakupan penelitian dibatasi pada halaman berbahasa Indonesia, sehingga halaman judi berbahasa lain berada di luar jangkauan sistem. Data dikumpulkan melalui *crawling* terhadap domain dari blocklist publik dan daftar situs Indonesia, menghasilkan 12.088 halaman mentah yang setelah pembersihan data menjadi 3646 halaman berlabel seimbang. Teks diolah melalui tahapan *Text Processing* meliputi *case folding*, pembersihan karakter, penyeragaman angka, penghapusan *stopword*, tokenisasi dengan kosakata 20.000 kata, dan penyeragaman panjang 300 token. Model klasifikasi dibangun dengan arsitektur Bidirectional Long Short-Term Memory (BiLSTM) dua lapis dan mekanisme Bahdanau Attention, dengan total 1.660.674 parameter. Pada 546 data uji, model mencapai *accuracy* 0,9762, *precision* 0,9780, *recall* 0,9745, F1-score 0,9762, dan ROC-AUC 0,9936. Model kemudian dipasang pada layanan lokal dan dihubungkan dengan ekstensi Google Chrome yang memblokir halaman berskor tinggi disertai kata pemicu sebagai bukti keputusan. Hasil ini menunjukkan penilaian isi teks mampu mengenali situs judi berbahasa Indonesia meskipun domainnya baru. Pengujian tambahan pada halaman berita bertopik judi memperlihatkan berita bercorak penegakan hukum dinilai aman, sedangkan berita bercorak edukasi bersudut pandang pemain masih dapat melampaui ambang, sehingga sistem dilengkapi pengaman domain media dan lembaga resmi.

Kata kunci: Bahdanau Attention; BiLSTM; deteksi judi online; ekstensi peramban; text processing

**REAL-TIME DETECTION OF ONLINE GAMBLING WEBSITES
BASED ON TEXT PROCESSING
USING THE LSTM ALGORITHM**

Zaidan Jibran Azhar, 202231116

Supervised by Dr. Ir. Abdul Haris, S.Kom., M.Kom., IPM.

ABSTRACT

Online gambling turnover in Indonesia reaches hundreds of trillions of rupiah per year, while address-based blocking quickly becomes outdated because gambling sites frequently change domains. This research designs and evaluates a real-time detection system for Indonesian-language online gambling websites that works inside the browser by assessing a page's textual content rather than its address. Pages in other languages fall outside its scope. Data were collected by crawling domains from public blocklists and Indonesian website lists, producing 12,088 raw pages that became a balanced labeled dataset of 3,646 pages after cleaning. The text passed through a Text Processing pipeline of case folding, character cleaning, number normalization, stopword removal, tokenization with a 20,000-word vocabulary, and padding or truncation to 300 tokens. The classification model used a two-layer Bidirectional Long Short-Term Memory (BiLSTM) architecture with a Bahdanau Attention mechanism, totalling 1,660,674 parameters. On 546 test pages, the model achieved an accuracy of 0.9762, a precision of 0.9780, a recall of 0.9745, an F1-score of 0.9762, and a ROC-AUC of 0.9936. The model was deployed on a local service connected to a Google Chrome extension that blocks high-scoring pages and presents trigger words as evidence. These results show that content-based assessment can recognize Indonesian-language gambling sites even on new domains. An additional test on news pages about gambling showed that law-enforcement reports score as safe, whereas educational articles written from a player's point of view may still exceed the blocking threshold, so a safeguard for registered press and official domains was added.

Keywords: Bahdanau Attention; BiLSTM; browser extension; online gambling detection; text processing