

I. JUDUL DAN ABSTRAK

1.1 Judul

BERTopic dan RoBERTa sebagai Model Analisis Keselarasan Peraturan Perundang-undangan tentang Investasi di Indonesia

1.2 Abstrak

Carut-marut peraturan perundangan terkait investasi di Indonesia disinyalir menjadi salah satu penghambat masuknya investor di Indonesia. Ego sektoral antar lembaga dan kepentingan daerah yang beragam juga dinilai menjadi penghambat pembangunan hukum di Indonesia. Ketidakseimbangan prioritas dan perspektif yang berbeda antar institusi dan daerah dapat menyulitkan dalam menyusun dan mengimplementasikan peraturan perundang-undangan yang konsisten dan efektif. Kondisi tersebut juga dipengaruhi dengan adanya *hyper-regulation*. Studi ini mencoba untuk mengkaji kesemrawutan peraturan perundangan dengan menggunakan BERTopic dan RoBERTa. Kedua model tersebut digunakan untuk menjawab permasalahan keselarasan atau keharmonisan peraturan perundang-undangan terkait investasi di Indonesia dengan menelaah 110 peraturan perundang-undangan terkait investasi. Hasil studi menunjukkan bahwa dari 110 peraturan perundang-undangan yang dianalisis hanya terdapat 0,01% pasangan ayat yang tidak selaras. Temuan studi ini menunjukkan bahwa dengan menggunakan BERTopic dan RoBERTa peraturan perundang-undangan terkait investasi di Indonesia sudah mendukung terciptanya lingkungan investasi yang efisien. Hasil studi ini dapat menjadi upaya awal untuk menganalisis peraturan perundang-undangan lainnya demi terciptanya konsistensi hukum di Indonesia.

Kata Kunci: BERTopic, Indonesia, Investasi, Peraturan perundang-undangan, RoBERTa

II. PENDAHULUAN

2.1 Latar Belakang

Pemerintah Indonesia memiliki kewajiban untuk melaksanakan pembangunan hukum nasional secara terencana, terpadu, dan berkelanjutan, dengan menjaga sistem hukum nasional yang menjamin perlindungan hak dan kewajiban seluruh rakyat Indonesia berdasarkan Undang-Undang Dasar Negara Republik Indonesia Tahun 1945 demi terwujudnya Indonesia sebagai negara hukum. Pembangunan hukum nasional tersebut dilakukan melalui pembentukan peraturan perundang-undangan beserta perangkatnya. Pembentukan peraturan perundang-undangan adalah proses pembuatan peraturan perundang-undangan yang mencakup tahapan perencanaan, persiapan, penyusunan, perumusan, pembahasan,

pengesahan, pengundangan, dan penyebarluasan produk hukum. Sementara itu, UU Nomor 12 Tahun 2011 dan pembaruan UU Nomor 15 Tahun 2019 menyebutkan bahwa peraturan-perundang-undangan merupakan norma tertulis yang memuat norma hukum yang mengikat secara umum dan dibentuk atau ditetapkan oleh lembaga negara atau pejabat yang berwenang melalui prosedur yang ditetapkan dalam peraturan perundang-undangan (Saputra, 2016).

Adanya penerapan otonomi daerah dan meluasnya kewenangan instansi lembaga membuat peraturan perundang-undangan di Indonesia semakin kompleks. Tidak sedikit terjadi tumpang tindih dan ketidaksesuaian perundang-undangan di Indonesia, terutama ketika terjadi transisi perubahan undang-undang. Ego sektoral antar lembaga dan kepentingan daerah yang beragam juga dinilai menjadi penghambat pembangunan hukum di Indonesia. Ketidakseimbangan prioritas dan perspektif yang berbeda antar institusi dan daerah dapat menyulitkan dalam menyusun dan mengimplementasikan peraturan perundang-undangan yang konsisten dan efektif.

Demikian halnya dengan peraturan perundang-undangan tentang investasi di Indonesia. Menteri Hukum dan HAM (Hak Asasi Manusia), Yasonna H. Laoly, pada peringatan Hari Pers Nasional 2020 menyatakan perlunya pembangunan hukum dan regulasi yang mendorong pertumbuhan ekonomi melalui investasi (Biro Humas, Hukum, dan Kerjasama Kemenkumham, 2020). Reformasi regulasi di bidang perizinan berusaha, ketenagakerjaan, pembangunan UMKM (Usaha Mikro Kecil dan Menengah), yang ditunjang dengan reformasi aturan perpajakan sangat penting dilakukan guna mencapai Program Kerja Prioritas Presiden, yang antara lain mencakup pembangunan sumber daya manusia dan penyederhanaan segala bentuk regulasi. Reformasi regulasi sangat penting untuk dilakukan guna menyelesaikan hambatan investasi, memotong rantai birokrasi yang panjang, serta untuk mengatasi tumpang-tindih dan aturan yang tidak harmonis (*hyper-regulation*).

Terkait dengan permasalahan tersebut di atas, kajian ini berusaha untuk menganalisis keselarasan peraturan perundangan tentang investasi di Indonesia dengan menerapkan model kajian *big data* dan *artificial intelligence*. *Big data* berpeluang dalam menganalisis peraturan perundang-undangan yang berada dalam posisi *hyper-regulation*. Selain itu, *artificial intelligence* membantu menyelesaikan permasalahan big data dari aspek komputasi dan algoritma. Salah satu alat yang populer untuk menyelesaikan permasalahan big data adalah LLM (*Large Language Model*).

2.2 Rumusan Masalah

Berdasarkan uraian latar belakang tersebut, kajian ini ingin melihat bagaimana tingkat keselarasan atau keharmonisan peraturan perundang-undangan di Indonesia terkait investasi dengan mengungkap permasalahan sebagai berikut.

1. Bagaimana pengaruh dari penggunaan BERTopic terhadap keefektifan dan efisiensi dalam analisis keselarasan peraturan perundang-undangan terkait investasi?
2. Bagaimana hasil prediksi dari model RoBERTa terhadap keselarasan peraturan perundang-undangan terkait investasi?
3. Bagaimana kinerja dan performa model RoBERTa dalam menganalisis keselarasan peraturan perundang-undangan terkait investasi?

2.3 Tujuan

Kajian tentang analisis keselarasan peraturan perundang-Undangn tentang investasi di Indonesia dengan menggunakan BERTopic dan RoBERTa memiliki tujuan sebagai berikut.

1. Untuk mengetahui bagaimana pengaruh dari penggunaan BERTopic terhadap keefektifan dan efisiensi analisis keselarasan peraturan perundang-undangan terkait investasi di Indonesia.
2. Menganalisis hasil prediksi model RoBERTa terhadap keselarasan peraturan perundang-undangan terkait investasi di Indonesia.
3. Untuk mengetahui kinerja atau performa model RoBERTa dalam menganalisis keselarasan peraturan perundang-undangan terkait investasi di Indonesia.

2.4 Manfaat

Kajian tentang analisis keselarasan peraturan perundang-Undangn tentang investasi di Indonesia dengan menggunakan BERTopic dan RoBERTa memiliki manfaat sebagai berikut.

1. Mencegah ambiguitas terhadap peraturan perundang-undangan terkait investasi di Indonesia.
2. Memberi masukan kepada pemerintah dalam mengevaluasi ayat berdasarkan hasil analisis keselarasan tiap ayat peraturan perundang-undangan sehingga dapat memperkuat kebijakan yang mendukung pertumbuhan investasi yang berkelanjutan dan berdampak positif bagi ekonomi dan masyarakat Indonesia.

3. Memberikan sumbangsih pemikiran terkait investasi dengan memastikan keselarasan peraturan perundang-undangan dan meminimalkan ketidakpastian hukum di Indonesia.

III. KAJIAN PUSTAKA

3.1 NLI (*Natural Language Inference*)

NLI (*Natural Language Inference*) adalah metode yang digunakan untuk menentukan hubungan antara hipotesis dengan premis yang diberikan. NLI mengklasifikasikan hipotesis menjadi tiga kelas berdasarkan premisnya, yaitu *entailment*, *contradiction*, dan *neutral*. Untuk melatih dan menguji model NLI, diperlukan dataset yang berisi pasangan premis dan hipotesis dengan label kelas yang sesuai (*entailment*, *contradiction*, atau *neutral*). Contoh dataset yang umumnya digunakan pada metode NLI adalah SNLI (*Stanford Natural Language Inference*) (Stanford NLP Group, n.d.) . Saat ini, terdapat beberapa model yang telah dilatih untuk tugas NLI seperti EFL (*Entailment as Few-Shot Learner*) (Wang et al., 2021), RoBERTa, dan ALBERT. Dengan kemampuan tersebut, NLI dapat dimanfaatkan untuk mengidentifikasi hubungan antara dua pernyataan.

3.2 BERTopic (*Bidirectional Encoder Representations from Transformers Topic*)

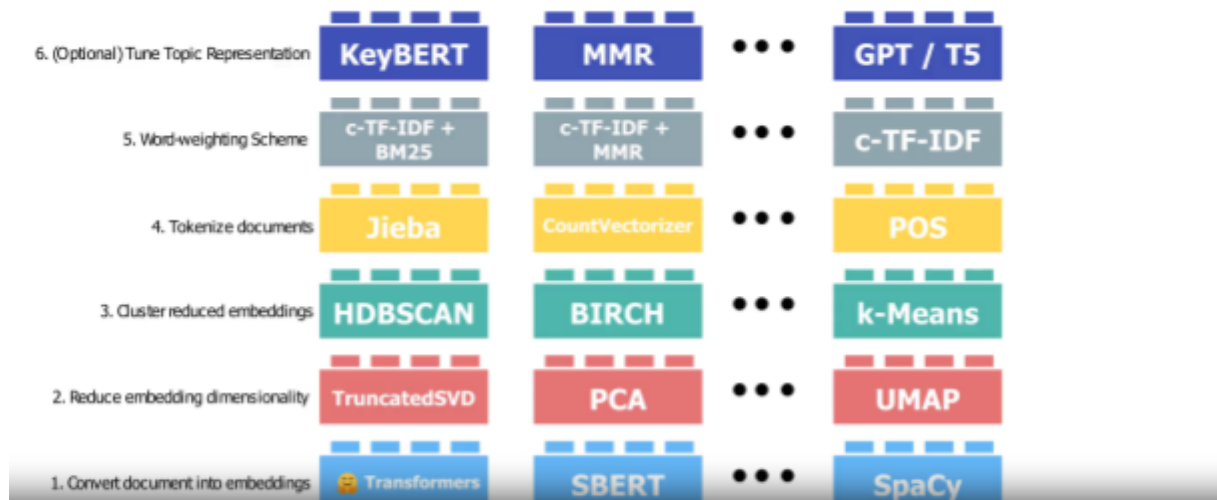
BERTopic adalah sebuah metode *topic modelling* yang menggunakan pendekatan berbasis *transformer*, yang utamanya menggunakan model bahasa BERT (*Bidirectional Encoder Representations from Transformers*). Metode ini dikembangkan untuk menemukan topik-topik atau pola tipikal yang ada pada koleksi dokumen atau teks tanpa supervisi.

Proses kerja BERTopic melibatkan langkah-langkah berikut (Grootendorst, 2022).

1. Proses BERTopic dimulai dengan menghasilkan representasi vektor dokumen menggunakan model bahasa berbasis *transformer* yang telah di-*pretrained*;
2. Setelah mengurangi dimensi representasi vektor dokumen tersebut, dilakukan *clustering* untuk membentuk kelompok dokumen yang memiliki kesamaan semantik, masing-masing merepresentasikan topik yang berbeda;
3. Terakhir, untuk mengatasi pandangan yang didasarkan pada *centroid*, BERTopic mengembangkan versi *class-based* model TF-IDF untuk mengekstrak representasi topik dari setiap kelompok dokumen.

Untuk melakukan semua proses tersebut, BERTopic menggunakan beberapa submodel yang dapat dikombinasikan. Setiap submodel yang digunakan memiliki enam kegunaan yaitu

embedding, reduksi dimensi, *clustering*, tokenisasi, *word weighing*, dan model opsional untuk melakukan *tuning* pada representasi topik.



Gambar 1. Submodel BERTopic
(Sumber: <https://maartengr.github.io/BERTopic/index.html>)

3.3 RoBERTa (*Robustly Optimized BERT Pre-training Approach*)

RoBERTa atau *Robustly Optimized BERT Approach* adalah sebuah model bahasa yang dikembangkan oleh FAIR (Facebook AI Research). Model ini merupakan varian dari BERT (*Bidirectional Encoder Representations from Transformers*) yang telah mengalami optimalisasi melalui penyesuaian pada teknik dan prosedur pada tahap *pre-training*. Menurut Y.Liu et al. (2019), BERT masih belum terlatih sepenuhnya dan dapat menyamai atau bahkan melebihi performa dari model yang dikeluarkan setelahnya. Model BERT merupakan model dengan arsitektur berbasis *Transformers* yang pada proses *pre-training*-nya dilakukan dengan menggunakan dua tugas, yaitu MLM (*Masking Language Modelling*) dan NSP (*Next Sentence Prediction*). Perbedaan proses dan teknik *pre training* yang dilakukan oleh model BERT dan RoBERTa adalah sebagai berikut (Liu et al., 2019).

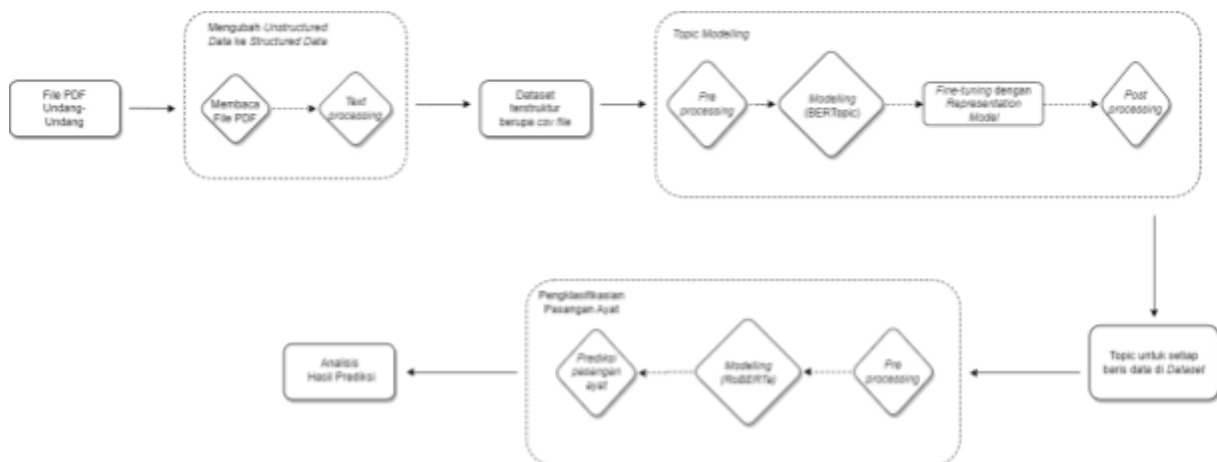
1. Pada model BERT proses *masking* hanya dilakukan pada data *preprocessing* sehingga menghasilkan *masking* yang bersifat statis. Hal ini mengakibatkan *input* yang sama akan berulang kali diberikan kepada model untuk setiap *epoch*-nya. Sementara pada model RoBERTa, proses *masking* bersifat lebih dinamis dimana data *training* akan diduplikasi sebanyak 10 kali dan di-*masking* dengan 10 cara yang berbeda sehingga mengurangi *input* yang sama pada model.

2. Pada model BERT *input* dari model berupa SEGMENT-PAIR + NSP (Next Sentence Prediction) Loss. Sementara pada model RoBERTa, *input* dari modelnya diubah menjadi FULL-SENTENCES serta tidak menggunakan NSP loss.
3. Model RoBERTa dilatih pada data *training* yang lebih banyak dibanding BERT serta dilatih pada *mini-batch* yang lebih besar dibanding model BERT.
4. Terkait *text encoding* model BERT menggunakan implementasi *character-level BPE* (Byte-Pair Encoding) *vocabulary*. Sementara itu, model RoBERTa menggunakan implementasi *byte-level BPE vocabulary*. Hal ini membuat ukuran *vocabulary* dari model RoBERTa menjadi lebih besar dibanding model BERT.

Optimalisasi tersebut membuat model RoBERTa memiliki performa yang lebih baik (sekitar 2-20%) dibanding model BERT (Singh, 2021) serta berhasil mendapatkan hasil *state-of-the-art* pada tiga benchmark berbeda, yaitu GLUE, SQuAD, dan RACE(Liu et al., 2019).

IV. METODOLOGI

Berikut adalah gambaran alur/proses penyelesaian masalah yang dilakukan.



Gambar 2. Diagram alir

4.1 Mengubah *Unstructured Data* ke *Structured Data*

Pertama, setiap file pdf akan dibaca dan diekstrak menggunakan *library* python PyMuPDF. *String* yang dihasilkan akan melalui tahap *processing* untuk mengekstrak setiap ayat pada peraturan perundang-undangan dan menghapus bagian-bagian yang tidak diperlukan. Proses *text processing* ini dilakukan dengan memperhatikan pola penulisan pada

file pdf dan dibantu dengan *regex* pada python. Selain itu, dilakukan penghapusan karakter *newline* dan spasi yang berlebih pada setiap ayat.

	NAMA_DOKUMEN	NO_PASAL	NO_AYAT	BUNYI_AYAT
11355	Peraturan_Presiden_No.9_Tahun_2011	1	2	Pinjaman dengan persyaratan lunak sebagaimana ...
11356	Peraturan_Presiden_No.9_Tahun_2011	2	1	Besarnya pinjaman dengan persyaratan lunak seb...
11357	Peraturan_Presiden_No.9_Tahun_2011	3	1	Pinjaman dengan persyaratan lunak kepada PT Pe...
11358	Peraturan_Presiden_No.9_Tahun_2011	4	1	Ketentuan lebih lanjut yang diperlukan dalam r...
11359	Peraturan_Presiden_No.9_Tahun_2011	5	1	Peraturan Presiden ini mulai berlaku pada tang...

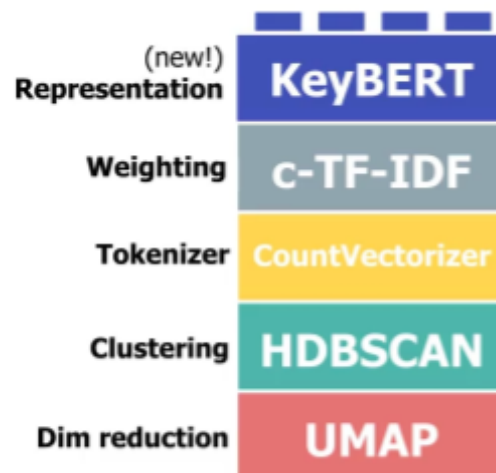
Gambar 3. *Structured Data*

4.2 Topic Modeling

Pada tahap ini, setiap ayat akan dikelompokkan berdasarkan topik yang dibahas pada ayat tersebut. Tahap ini bertujuan agar proses pengidentifikasian keselarasan antara setiap pasang ayat menjadi efisien. Dengan melakukan *topic modeling*, pasangan ayat yang perlu diidentifikasi keselarasan ayatnya hanya ayat yang membahas topik yang sama. Sementara pasangan ayat yang berbeda topiknya tidak perlu untuk diidentifikasi karena kedua pasangan ayat tersebut pasti saling netral sehingga tergolong pada hubungan ayat yang selaras.

Berikut adalah proses yang dilakukan pada tahap *topic modeling*.

1. Setiap ayat pada *dataset* akan dilakukan proses *embedding*. Proses *embedding* berguna untuk mengubah input menjadi representasi numerik. Proses ini dilakukan menggunakan *Sentence Transformers* dengan model yang digunakan adalah *paraphrase-multilingual-MiniLM-L12-v2* yang mendukung banyak bahasa termasuk bahasa Indonesia dan memiliki ukuran yang lebih kecil dibanding model lain.
2. Pemodelan model yang digunakan untuk melakukan *clustering* topik pada setiap ayat. Pada tahap ini digunakan teknik *topic modeling* BERTopic, dimana untuk *sub* model yang digunakan adalah UMAP sebagai *dimensionality reduction*, HDBSCAN sebagai *clustering model*, CountVectorizer sebagai *tokenizer*, dan c-TF-IDF sebagai skema *word weighting*. Pada model juga ditetapkan parameter *min_topic_size* menjadi 2 sehingga suatu topik dapat dibentuk jika terdapat 2 ayat yang memiliki *cluster* yang sama. Hal ini dilakukan agar *outliers* menjadi lebih sedikit sehingga mengurangi data yang hilang.



Gambar 4. Struktur Model

3. *Fine-tuning keywords* yang dihasilkan oleh c-TF-IDF menggunakan model KeyBERT. Dengan menerapkan hal ini, representasi topik memiliki hasil yang lebih koheren dan mampu mengurangi adanya kata henti atau *stopwords* pada representasi topik yang dihasilkan.
4. Hasil dari proses sebelumnya akan dilakukan *post processing*. *Post processing* yang dilakukan bertujuan untuk mengurangi *mini cluster* sekaligus mengurangi *outliers* pada dataset sehingga dapat mengurangi jumlah data yang hilang. Proses ini dilakukan dengan menggabungkan topik-topik yang mirip dari semua topik yang sebelumnya dihasilkan model.

4.3 Pengklasifikasian Pasangan Ayat

Pada tahap ini, akan dilakukan klasifikasi pada setiap pasangan ayat yang memiliki topik yang sama. Sebelumnya, diperlukan tahap *pre-processing* pada dataset untuk mempersiapkan data sebelum dijalankan dalam algoritma klasifikasi. Tahap *pre-processing* dilakukan dengan melakukan pengelompokkan baris data berdasarkan kategori topik yang dimiliki. Pada setiap kelompoknya, setiap baris data akan dipasangkan dengan baris data lain menggunakan library *itertools* pada python. Terakhir, setiap pasangan ayat akan digabungkan pada satu *dataframe*

	NAMA_DOKUMEN	NO_PASAL	NO_AYAT	BUNYI_AYAT	NAMA_DOKUMEN_LAIN	NO_PASAL_LAIN	NO_AYAT_LAIN	BUNYI_AYAT_LAIN	TOPIC
0	2022perpu002	51	15	Rencana fungsi dan pemanfaatan sebagaimana dim...	2022perpu002	51	26	Pemerintah Daerah menerbitkan sertifikat lak ...	15
1	Undang-Undang_No.8_Tahun_1995	105	1	Manajer Investasi dan atau Pihak terafiliasiny...	Undang-Undang_No.40_Tahun_2014	75	1	Setiap Orang yang dengan sengaja tidak memberl...	0
2	2022perpu002	175	17	Ketentuan mengenai jenis, bentuk, dan mekanism...	Perpu Nomor 1 Tahun 2020	5	3	Ketentuan lebih lanjut mengenai persyaratan te...	5
3	2022perpu002	17	105	Jika tindak pidana sebagaimana dimaksud pada a...	2022perpu002	46	97	Setiap Pelaku Usaha yang tidak menggunakan ata...	0

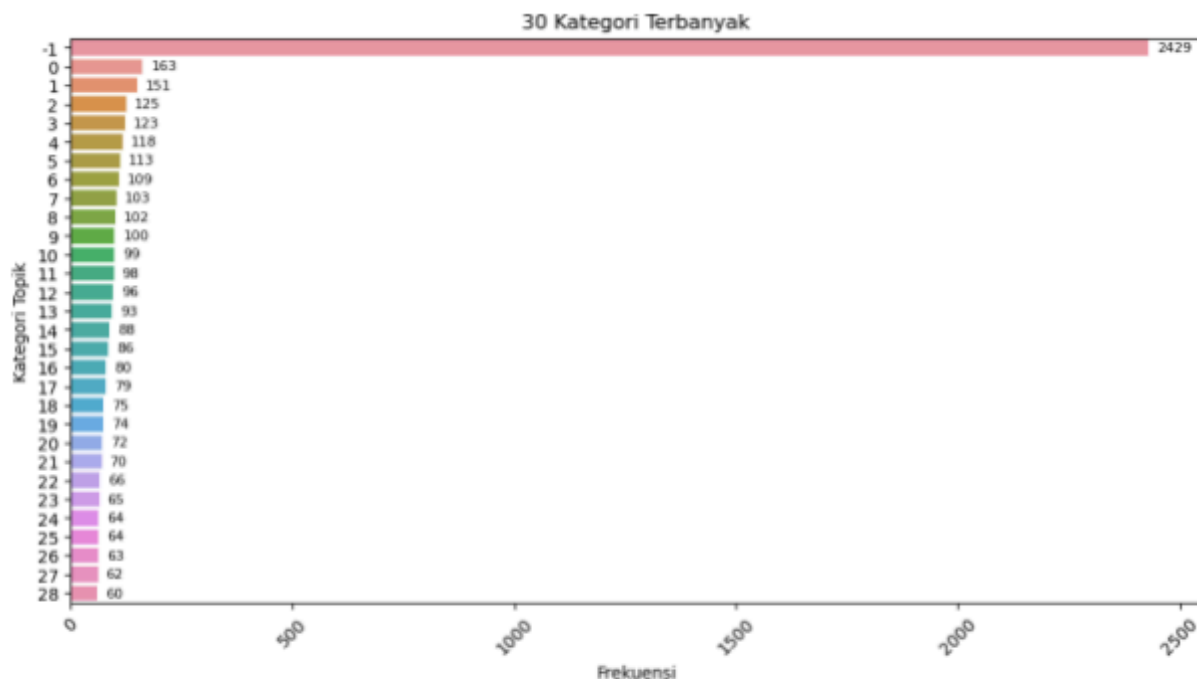
Gambar 5. Hasil *pre-processing*

Untuk melakukan klasifikasi pada setiap pasangan ayat di setiap baris datanya digunakan *pre-trained* model Indo-roberta-indonli (Limcorn, 2022). Indo-roberta-indonli sendiri adalah model klasifikasi NLI (*natural language inference*) yang didasari oleh model Indonesian RoBERTa (Wongso et al., 2021) yang merupakan versi bahasa Indonesia dari model RoBERTa. Model Indo-roberta-indonli telah dilatih pada dataset IndoNLI (Mahendra et al., 2021) sehingga memungkinkan model untuk mengklasifikasikan dua pernyataan/argumen pada kategori *entailment*, *neutral*, dan *contradiction*. Model ini dapat digunakan untuk menganalisis hubungan antara dua ayat pada peraturan perundang-undangan dimana untuk pasangan ayat yang memiliki hubungan bersifat *entailment* atau *neutral* maka pasangan ayat tersebut tergolong pada pasangan ayat yang selaras. Sementara pasangan ayat yang memiliki hubungan *contradiction* tergolong pada pasangan ayat yang tidak selaras.

V. PEMBAHASAN

Pada awalnya terdapat 110 dokumen peraturan perundang-undangan yang terdiri dari dari Peraturan Pemerintah, Peraturan Presiden, Peraturan Menteri, Undang-undang, dan Peraturan Pemerintah Pengganti Undang-Undang. Pengubahan *unstructured data* menjadi *structured data* menghasilkan jumlah ayat sebanyak 11360 ayat. Artinya akan dihasilkan sebanyak 64.519.120 pasang ayat yang perlu diidentifikasi keselarasannya. Dengan adanya *topic modeling*, pasangan ayat yang perlu diidentifikasi keselarasannya berjumlah 178.845 pasang ayat sehingga proses analisis menjadi lebih efisien.

Pada tahap *topic modeling* dihasilkan sebanyak 949 kategori topik dengan frekuensi dari 30 kategori terbanyak dapat dilihat pada grafik berikut ini.



Gambar 6. Grafik frekuensi topik

Pada grafik di atas, kategori topik bernomor “-1” merupakan *outliers* dimana tidak terdapat ayat lain yang memiliki kategori yang sama dengan ayat tersebut. Sementara kategori topik bernomor “0” merupakan topik dengan frekuensi terbanyak sebanyak 163 ayat.

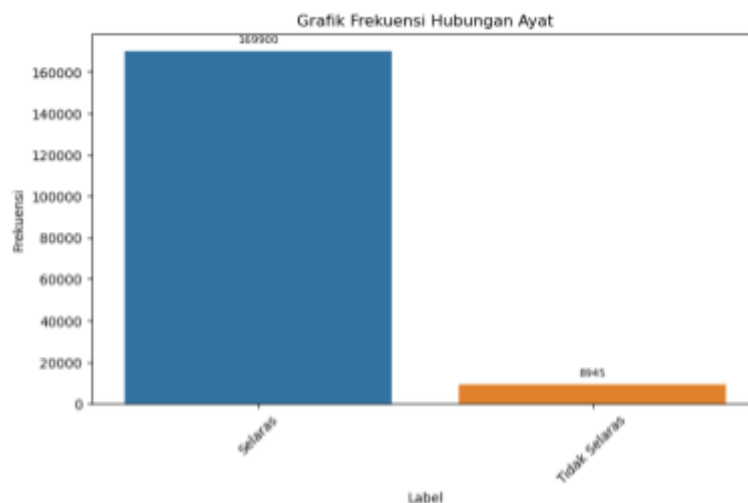
Untuk representasi *keyword* pada beberapa topik yang dihasilkan model dapat dilihat pada grafik dibawah ini.



Gambar 7. Topic word scores

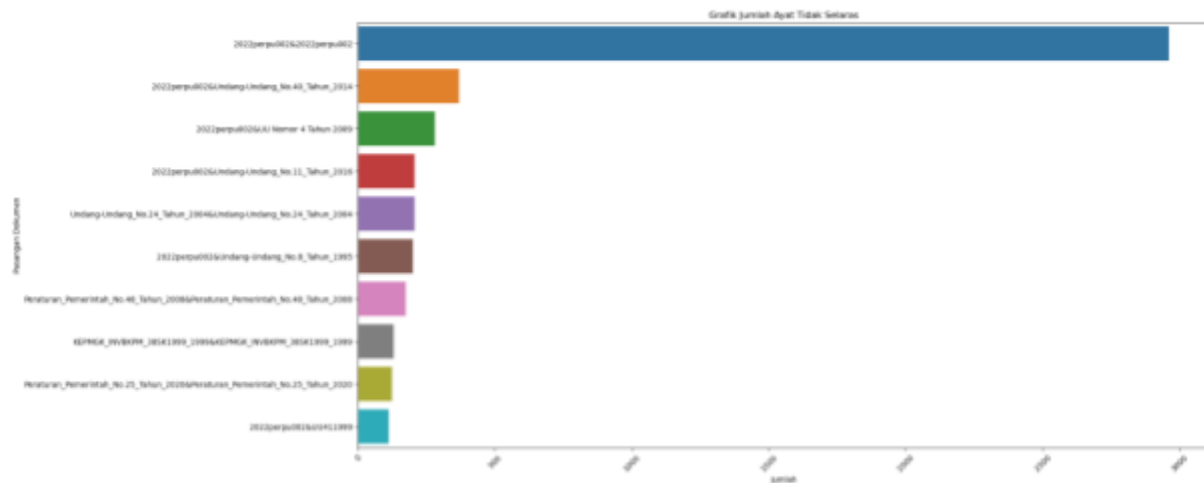
Sumbu x pada grafik di atas menggambarkan skor dari c-TF-IDF yang telah di *fine tuning* dengan KeyBERT. Sementara sumbu y menggambarkan 5 kata atau *keyword* dengan skor tertinggi yang merepresentasikan topik terkait. Berdasarkan grafik tersebut terlihat bahwa kebanyakan ayat pada peraturan perundang-undangan yang diberikan membahas terkait hukuman pidana atau penjara dengan kata kuncinya yaitu penjara, pidana, miliar, juta, dan korban.

Berdasarkan prediksi yang dihasilkan oleh model, sebanyak 169.900 pasang ayat atau sekitar 95% dari pasangan ayat yang dianalisis memiliki hubungan yang saling selaras. Jika dihitung persentasenya terhadap semua pasangan ayat yang ada yaitu 64.519.120 pasang ayat, maka persentase pasangan ayat yang tidak selaras hanya berkisar 0,01% saja.



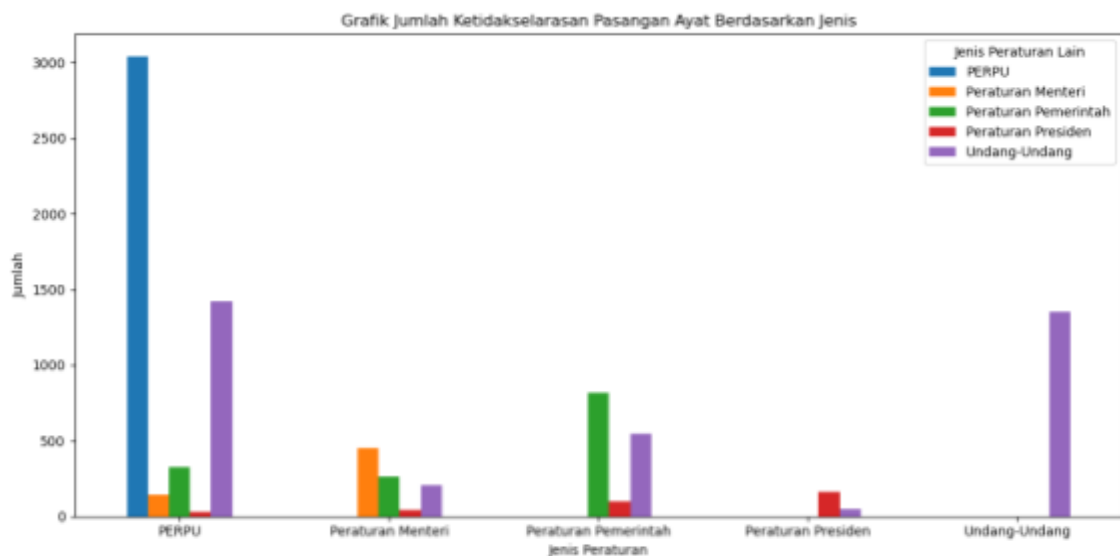
Gambar 8. Grafik Prediksi Model

Sementara itu, jika pasangan ayat yang memiliki hubungan tidak selaras dianalisis lebih lanjut, ditemukan bahwa pasangan ayat pada Peraturan Pemerintah Pengganti Undang-Undang Nomor 2 Tahun 2022 memiliki pasangan ayat tidak selaras paling banyak yaitu sebanyak 2961 pasang.



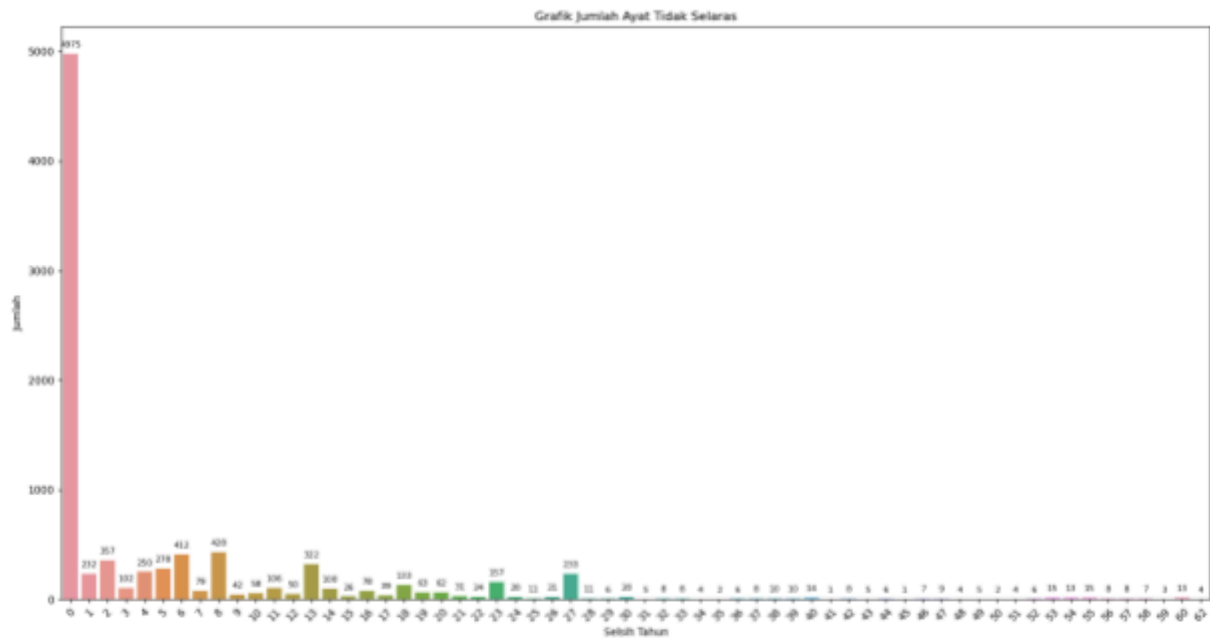
Gambar 9. Grafik Jumlah Ayat Tidak Selaras

Jika dilihat berdasarkan jenisnya dari peraturan perundang-undangannya, terlihat bahwa model cenderung menemukan pasangan-pasangan ayat yang tidak selaras pada jenis peraturan perundang-undangan yang sama.



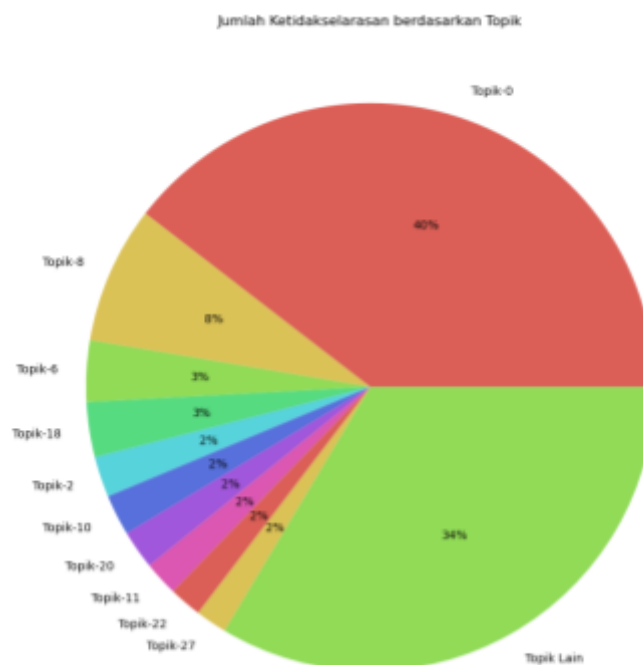
Gambar 10. Jumlah Ketidakselarasan Berdasarkan Jenis

Pada pasangan-pasangan ayat yang tidak selaras terlihat juga bahwa keduanya cenderung disahkan pada tahun yang sama, seperti yang terlihat pada grafik berikut.



Gambar 11. Selisih Tahun pada Pasangan Ayat Tidak Selaras

Berdasarkan topiknya, pasangan ayat yang tidak selaras cenderung memiliki topik nomor 0 pada **Gambar 11**, yaitu membahas hukuman pidana atau penjara dimana sekitar 40% dari total pasangan ayat yang tidak selaras.



Gambar 12. Jumlah Ayat Tidak Selaras Berdasarkan Topik

Pada dataset yang berisi 1057 baris data yang telah di-*labeling* secara manual, model yang digunakan memiliki akurasi sebesar 93.34%. Akan tetapi, masih cukup banyak *false positive* pada prediksi yang dihasilkan oleh model dimana kalimat yang seharusnya tidak mengandung kontradiksi, tetapi terdeteksi mengandung kontradiksi. Hal ini cenderung terjadi pada dokumen peraturan perundang-undangan yang kompleks, salah satunya Peraturan Pemerintah Pengganti Undang-Undang Nomor 2 Tahun 2020 yang memiliki lebih dari 1000 halaman yang mengakibatkan model kesulitan memahami konteks pada peraturan perundang-undangan tersebut. Untuk itu, model tersebut dapat dikembangkan lagi dengan melatihnya pada dataset yang mengandung pernyataan-pernyataan di dalam peraturan perundang-undangan. Hal ini dapat membuat model menjadi lebih baik dalam memahami dan menganalisis kalimat-kalimat pada undang-undang.

VI. PENUTUP

6.1 Kesimpulan

Peraturan perundang-undangan mengenai investasi cenderung memiliki tingkat keharmonisan atau keselarasan yang baik karena hanya sebagian kecil dari pasangan ayat yang mengandung kontradiksi. Model Indo-roberta-indonli juga memiliki performa yang cukup baik dalam mengidentifikasi keselarasan pada peraturan perundangan-undangan terkait investasi. Akan tetapi, model tersebut memiliki performa yang tidak terlalu baik pada dokumen perundang-undangan yang kompleks. Selain itu, penggunaan BERTopic juga terbukti membantu proses analisis keselarasan peraturan perundang-undangan menjadi lebih efisien.

6.2 Saran

Sebagai negara hukum, keharmonisan atau keselarasan undang-undang adalah sebuah hal yang perlu untuk diperhatikan dan dianalisis dengan seksama. Analisis keharmonisan undang-undang menjadi sangat penting karena undang-undang merupakan pijakan hukum yang mengatur kehidupan berbagai lapisan masyarakat. Untuk itu, Indonesia sebaiknya dapat menyediakan dan menyimpan data peraturan perundangan-undangan dengan lebih baik serta melibatkan orang-orang yang ahli dalam bidang hukum dalam pengumpulan ataupun *labeling* data. Dengan adanya data tersebut, dapat dibuat sistem atau model yang lebih akurat dalam menganalisis keselarasan peraturan perundang-undangan sehingga mengurangi ketidakpastian hukum di Indonesia.

VII. REFERENSI

Biro Humas, Hukum, dan Kerjasama Kemenkumham. (2020). *Menkumham: Kemudahan Investasi Demi Kemakmuran Masyarakat Indonesia*.
<https://www.kemenkumham.go.id/berita-utama/menkumham-kemudahan-investasi-demi-kemakmuran-masyarakat-indonesia>

Saputra, A. (2016, September 29). *Paket Reformasi Hukum, Jokowi Harus Tuntaskan Egosektoral Antar Lembaga*. detikNews.
<https://news.detik.com/berita/d-5995527/peraturan-perundang-undangan-pengertian-jenis-hingga-muatan-materi>

Stanford NLP Group. (n.d.). *The Stanford Natural Language Inference (SNLI) Corpus*.
<https://nlp.stanford.edu/projects/snli/>

Wang, S., Fang, H., Khabsa, M., Mao, H., Ma, H. (2021). *Entailment as Few-Shot Learner*. <https://arxiv.org/pdf/2104.14690>

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. <https://arxiv.org/abs/2203.05794>

Singh, A. (2021). *Evolving with BERT: Introduction to RoBERTa*.
<https://medium.com/analytics-vidhya/evolving-with-bert-introduction-to-roberta-5174ec0e7c82>

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. <https://arxiv.org/abs/1907.11692>

Limcorn, S. (2021). *Indo-roberta-indonli*. HuggingFace.
<https://huggingface.co/StevenLimcorn/indo-roberta-indonli>

Wongso, W., Limcorn, S., Rahmadani, S., Wah, C.K. (2021). *Indonesian RoBERTa Base*. HuggingFace. <https://huggingface.co/flax-community/indonesian-roberta-base>

Mahendra, Rahmad and Aji, Fikri, A. and Louvan, Samuel and Rahman, Fahrurrozi and Vania, Clara. (2021, December 5). *IndoNLI: A Natural Language Inference Dataset for Indonesian*. <https://github.com/ir-nlp-csui/indonli/tree/main>