BAB I

PENDAHULUAN

1.1. Latar Belakang

Berkembangnya ilmu teknologi merupakan perkembangan yang bersifat *rapid* dan tidak pernah stagnan. Setiap disiplin ilmu dan pelajar atau mahasiswa harus mengikuti perkembangan ini dalam rangka tidak tertinggal dari perkembangan teknologi yang ada. Salah satu upaya dalam hal ini adalah dengan mengikuti perlombaan atau turnamen dalam suatu bidang. Dalam kegiatan ini, secara lebih khusus adalah turnamen dalam bidang sains data dalam skala nasional.

Turnamen sains data diadakan untuk mendorong inovasi dan kreativitas dalam pengolahan dan analisis data. Dengan menantang peserta dengan masalah dunia nyata yang kompleks, turnamen ini merangsang pengembangan pemecahan masalah yang mendalam dan meningkatkan kualitas model prediktif serta metode analisis data [1]. Selain itu, kompetisi ini menjadi wadah untuk berbagi pengetahuan dan teknik terbaik di antara komunitas data sains, memungkinkan pertukaran ide yang dapat menghasilkan solusi yang lebih baik. Turnamen sains data juga menjadi *platform* rekrutmen bakat bagi perusahaan yang mencari individu dengan keahlian yang mendalam dalam analisis data, sementara bagi peserta, kompetisi ini meningkatkan profil mereka dan memberikan kesempatan untuk mendemonstrasikan kemampuan mereka di tingkat global [2]. Dengan demikian, turnamen sains data tidak hanya mengukur kemampuan individu, tetapi juga berkontribusi pada pengembangan lapangan data sains secara keseluruhan. Dalam hal ini, Data Academy menyelenggarakan turnamen sains data yang dinamakan dengan Turnamen Sains Data Nasional (TSDN).

TSDN merupakan hasil kerjasama antara beberapa perguruan tinggi, organisasi data di Indonesia, pemerintah daerah, berbagai lembaga negara, dan

kementerian, yang dipimpin oleh Data Academy. Luthfy Ardiansyah, selaku Direktur Data Academy dan koordinator penyelenggara Turnamen Sains Data Nasional 2023, menyatakan dalam pernyataan tertulisnya bahwa harapannya adalah TSDN 2023 dapat diikuti oleh peserta yang lebih banyak dan bervariasi, termasuk pelajar SMA/SMK, mahasiswa, dan para profesional yang berkecimpung di bidang data sains. Ambisi Data Academy adalah agar acara ini dapat memicu minat dan menumbuhkan bakat baru di bidang sains data, terutama di kalangan generasi muda. Luthfy Ardiansyah menegaskan dalam keterangannya bahwa TSDN 2023 diharapkan mampu menciptakan persaingan inovatif dan solutif yang lebih mendalam, dengan menyoroti relevansi ilmu sains data dalam kehidupan sehari-hari, khususnya dalam meningkatkan keamanan *cyber* melalui pendekatan data sains dan kecerdasan buatan (AI). Fokus TSDN tahun ini adalah pada pengembangan pemahaman tentang metode, pendekatan, dan cara berpikir yang mendukung para ahli keamanan data. Peserta diharapkan dapat menghadapi berbagai tantangan seputar analisis keamanan menggunakan data sains dan teknik AI untuk mencapai solusi terbaik. Dalam proses penilaian, ahli sains data dan akademisi akan bertindak sebagai juri yang mengevaluasi solusi peserta berdasarkan kreativitas, akurasi, dan keteraplikasiannya di dunia nyata. Tak hanya itu, peserta juga diberikan peluang untuk membangun jaringan dengan para profesional dan rekan sesama penggemar sains, mendorong kolaborasi dan pertukaran pengetahuan [3].

Dalam evolusi teknologi, peran media sosial, seperti contohnya media sosial X, memiliki dampak signifikan. X, yang mengubah namanya dari Twitter sejak tahun 2022, merupakan *platform* jejaring sosial *online* terkemuka yang telah menjadi arena kolaborasi dan komunikasi bagi jutaan pengguna sejak diluncurkan pada tahun 2006. Pada tahun 2018, dengan lebih dari 320 juta pengguna aktif bulanan, X digunakan untuk berbagai keperluan, termasuk pengiriman pesan singkat atau "*tweet*" dengan batasan 140 karakter. Selain itu, pengguna dapat berbagi konten multimedia, termasuk video, sebagai bentuk komunikasi di platform ini [4] . Pertumbuhan pesat penggunaan X, didorong

oleh interaksi sosial sehari-hari antar pengguna, juga menarik perhatian penjahat dunia maya yang memanfaatkan hubungan kepercayaan antar pengguna untuk menyebarkan konten berbahaya kepada sejumlah besar korban di jaringan. Penjahat dunia maya di X, yang sering disebut sebagai *spammer* sosial atau akun yang mencurigakan, bekerja sama dengan berbagai cara untuk meningkatkan visibilitas akun mereka [4]. Mereka memanfaatkan X untuk mengirimkan pesan kekerasan, tautan *phishing*, perundungan siber, dan kejahatan siber lainnya. Selain itu, akun spam ini juga menyebarkan berita palsu yang dianggap sebagai ancaman serius terhadap demokrasi, jurnalisme, dan perekonomian [5].

Kelompok pelaku spam ini mengandalkan layanan yang disediakan oleh X, termasuk penggunaan simbol "#" untuk mengelompokkan *tweet* berdasarkan topik. Dengan cepatnya perkembangan berita palsu dan dampak negatifnya terhadap demokrasi, keadilan, serta tingkat kepercayaan masyarakat, terjadi peningkatan permintaan akan deteksi berita palsu [6] . Penyebaran berita palsu saat ini sangat cepat, dan dampaknya dapat menyebabkan pengetahuan yang keliru. Melalui akun spam, penyerang dapat merubah lingkungan Jaringan Sosial *Online* (OSN) yang dapat dipercaya melawan penggunanya dengan menyebarkan pesan spam di jaringan tersebut. Meningkatnya serangan berbahaya di situs jejaring sosial secara eksponensial menuntut kebutuhan akan solusi untuk mencegah serangan ini dan melindungi pengguna dari penipuan [7]. Oleh karena itu, diperlukan suatu kerangka kerja yang komprehensif untuk secara konsisten memahami dan mengidentifikasi berita palsu, dengan tujuan untuk mengumpulkan dan menyatukan peneliti dari berbagai latar belakang dalam upaya penelitian terkait berita palsu [7].

Penelitian dilakukan dengan mengembangkan *unsupervised machine learning* menggunakan PCA dan K-means, namun sebelum itu, dilakukan proses SNA. Analisis Jaringan Sosial atau *Social Network Analysis* (SNA) adalah suatu metode analisis yang melibatkan identifikasi dan pemahaman

interaksi serta hubungan antara entitas dalam suatu jaringan sosial [8]. Dalam konteks profesional, SNA memiliki komponen utama seperti *node*, *edge*, dan graf yang menciptakan representasi visual atau matematis dari jaringan [9]. Pengukuran sentralitas yang dilakukan pada penelitian ini meliputi *degree centrality*, *closeness centrality*, dan *eigenvector centrality*, serta *betweenness centrality* membantu mengidentifikasi entitas yang berperan penting dalam jaringan. *Degree centrality* adalah salah satu metrik sentralitas dalam analisis jaringan yang digunakan untuk mengukur sejauh mana suatu simpul (*node*) dalam jaringan memiliki koneksi dengan simpul-simpul lainnya. Metrik ini mencerminkan seberapa banyak hubungan yang dimiliki oleh suatu simpul [10]. *Closeness centrality* adalah salah satu metrik sentralitas dalam analisis jaringan yang mengukur seberapa dekat suatu simpul dengan simpul-simpul lain dalam jaringan. Metrik ini mencerminkan seberapa cepat suatu simpul dapat dijangkau atau diakses oleh simpul-simpul lain dalam jaringan. *Closeness centrality* didefinisikan sebagai invers dari jumlah rata-rata jarak terpendek antara simpul tertentu dengan semua simpul lain dalam jaringan [11]. *Betweenness centrality* adalah metrik sentralitas dalam analisis jaringan yang mengukur sejauh mana suatu simpul berada di antara jalur terpendek antara pasangan simpul lain dalam jaringan. Simpul dengan *betweenness centrality* yang tinggi cenderung berada di jalur-jalur kritis dalam jaringan dan dapat memainkan peran penting dalam menghubungkan bagian-bagian yang berbeda atau mengontrol aliran informasi [12]. *Eigenvector centrality* adalah metrik sentralitas dalam analisis jaringan yang memberikan bobot atau skor kepentingan pada simpul dalam jaringan berdasarkan hubungannya dengan simpul-simpul penting lainnya. Sebuah simpul memiliki *eigenvector centrality* yang tinggi jika ia terhubung dengan simpul-simpul yang juga memiliki *eigenvector centrality* yang tinggi [13]. Contoh penerapan SNA meliputi analisis interaksi karyawan dalam sebuah organisasi atau penyebaran informasi di media sosial seperti Twitter. Dalam penerapan *Machine Learning* (ML), informasi dari SNA dapat meningkatkan kinerja model dengan mendukung fungsi seperti prediksi hubungan,

pengelompokan, rekomendasi, dan deteksi anomali [14]. Dengan demikian, integrasi SNA dan ML memberikan wawasan mendalam tentang struktur jaringan dan dinamika hubungan, memperkaya pemahaman tentang interaksi dalam suatu konteks tertentu.

Dengan demikian, melalui pemaparan pentingnya berpartisipasi dalam sebuah turnamen sains data serta pemahaman esensi dari bahayanya akun yang dapat menyebarkan berita hoaks, *cyber crime*, *cyber bullying*, *phising,* dan dampak negatif lainnya, maka laporan ini disusun untuk dokumentasi sebagai partisipan TSDN 2023 dengan tema "Meningkatkan Keamanan Menggunakan Keilmuan Sains Data dan AI" dengan mengimplementasikan *Social Network Analysis* terhadap data yang kemudian nantinya akan diimplementasikan oleh machine learning dalam mendeteksi *suspicious account*.

1.2. Rumusan Masalah

Dari pemaparan latar belakang di atas, maka laporan ini mencari jawaban atas rumusan masalah sebagai berikut:

1. Bagaimana proses SNA (*Social Network Analysis*) untuk *feature extraction* terhadap deteksi *suspicious account* yang meliputi *degree centrality, closeness centrality, betweness centrality,* dan *eigenvector centrality*?
2. Bagaimana cara melakukan visualisasi dari SNA menggunakan *tools* Gephi?

1.3. Maksud dan Tujuan (tujuan penelitian)

Adapun maksud dalam melaksanakan perlombaan Turnamen Sains Data Nasional 2023 yang diselenggarakan oleh Data Academy adalah di antaranya sebagai berikut:

a. Pengembangan keterampilan praktis.

Mengikuti turnamen ini memiliki tujuan untuk mengembangkan dan meningkatkan keterampilan praktis dalam analisis data, pengolahan data, dan penggunaan alat dan teknik data sains.

b. Pengimplementasian pengetahuan teoritis.

Menerapkan pengetahuan teoritis yang diperoleh dari pembelajaran di kelas ke dalam konteks dunia nyata, dengan tujuan untuk meningkatkan pemahaman dan penerapan konsep akademis.

c. Pengalaman kompetisi.

Mengalami suasana kompetisi di bidang data sains yang dapat memberikan dorongan motivasi tambahan dan kesempatan untuk berinteraksi dengan rekan-rekan yang memiliki minat serupa.

d. Pengenalan pada industri.

Memahami bagaimana pemrosesan dan analisis data digunakan di industri, serta mendapatkan pandangan langsung terhadap tantangan dan tren terkini dalam domain data sains.

e. *Networking* dan kesempatan yang luas.

Membangun jaringan dengan profesional dan perusahaan yang berkecimpung dalam bidang data sains, memberikan kesempatan untuk mendapatkan wawasan industri dan juga peluang kerja.

f. Peningkatan profil akademis dan profesional.

Meraih prestasi dalam perlombaan dapat meningkatkan profil akademis dan profesional mahasiswa, memberikan keuntungan kompetitif di pasar kerja atau dalam melanjutkan studi.

g. Pengembangan proyek independen.

Melibatkan diri dalam perlombaan data sains memberikan kesempatan untuk mengembangkan proyek independen, memahami proses dari awal hingga akhir, dan mengasah kemampuan manajemen proyek.

h. Keterlibatan dalam komunitas data sains.

UNIVERSITAS MULTIMEDIA NUSANTARA

Mahasiswa dapat memiliki tujuan untuk aktif terlibat dalam komunitas data sains, berkontribusi pada solusi masalah di dunia nyata dan memperoleh apresiasi dari para profesional di industri.

i. Pengembangan portofolio.

Membangun portofolio proyek data sains yang sukses, yang dapat dipresentasikan kepada calon majikan atau institusi pendidikan yang tertarik.

j. Mendukung tujuan akademis.

Melibatkan diri dalam perlombaan data sains dapat mendukung tujuan akademis mahasiswa, seperti penelitian atau studi lanjutan dalam bidang data sains atau ilmu terkait.

k. Menjadi program MBKM dan pengajuan kelanjutan skripsi.

MBKM atau Mata Kuliah Berkehidupan Mahasiswa adalah inisiatif yang diperkenalkan oleh Kementerian Pendidikan dan Kebudayaan Republik Indonesia. Program ini dirancang untuk memberikan pengetahuan, keterampilan, dan sikap positif yang relevan dengan kehidupan sehari-hari dan persiapan menghadapi dunia kerja bagi mahasiswa. MBKM dirancang agar dapat diintegrasikan ke dalam kurikulum program studi dan memberikan bobot kredit yang dapat diakumulasi selama masa kuliah.

Di samping itu, tujuan dilaksanakannya TSDN 2023 oleh Data Academy ini adalah diharapkan mampu memberikan dampak positif yang signifikan pada kemajuan teknologi di Indonesia, sekaligus melahirkan talenta mumpuni dalam bidang *data science* dan kecerdasan buatan (AI). Keberhasilan turnamen diharapkan dapat menciptakan inovasi-inovasi baru yang dapat diterapkan dalam meningkatkan sistem keamanan siber, menghadirkan solusi yang efektif dan canggih berbasis ilmu data dan kecerdasan buatan untuk mengatasi tantangan keamanan siber yang semakin kompleks. Hal ini direalisasikan melalui melaksanakan beberapa tahapan, prosedur, syarat dan ketentuan, hingga projek dapat dirampungkan.

1.4. Manfaat

Manfaat yang diperoleh dari partisipasi dalam TSDN 2023 dapat dibagi menjadi dua kategori, yakni manfaat akademis dan manfaat praktis.

a. Manfaat akademis, yaitu mendapatkan penerapan teori dalam konteks nyata, peningkatan pemahaman algoritma dan metodologi, dan pembelajaran secara kolaboratif.
b. Manfaat praktis, yaitu mendapatkan pengalaman yang relevan dan nyata serta pengembangan keterampilan teknis dan praktis.

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA