

BAB III

METODE PELAKSANAAN

3.1 Tahap Pelaksanaan Program

3.1.1 Waktu dan Prosedur Pelaksanaan Proyek Independen

3.1.1.1 Waktu Pelaksanaan Kerja Proyek Independen

Periode kerja magang dimulai pada tanggal 1 Agustus 2023 dan berakhir pada tanggal 19 November 2023. Waktu kerja dapat disesuaikan secara fleksibel setiap hari, dan laporan harian tugas dilakukan melalui situs web www.merdeka.umn.ac.id. Komunikasi dengan pihak Data Academy dilakukan secara daring melalui aplikasi WhatsApp, sementara komunikasi dengan *supervisor* lapangan dilakukan dalam mode *hybrid*, yakni baik secara tatap muka maupun daring.

1.3.1.2 Prosedur Pelaksanaan Proyek Independen

Sebagai bagian dari pelaksanaan TSDN 2023 di Data Academy, proses ini melewati beberapa tahapan atau langkah-langkah tertentu. Berikut adalah beberapa langkah yang perlu diikuti dalam pelaksanaan tersebut:

1. Pra-Turnamen

Dilakukan dengan mencari informasi terkait kompetisi-kompetisi yang bergerak di bidang informasi teknologi dengan skala nasional. Pencarian ini dilakukan melalui *browsing* secara independen di media sosial dan Google serta pemberian informasi melalui dosen dan himpunan mahasiswa.

2. Turnamen

Dilakukan dengan mengikuti seluruh rangkaian kegiatan turnamen yang dimulai dari Pendaftaran TSDN (10-28 Oktober 2023), Pembukaan TSDN (28 Oktober 2023), Pelatihan: *Sharing Session & Workshop* (30 Oktober-4 November 2023), Batas Akhir Pengumpulan Proposal (8-16 November 2023),

Babak Penyisihan (17-23 November 2023) hingga Pengumuman Finalis (29 November 2023).

3. **Pasca Turnamen**

Pasca turnamen dilakukan melalui penyusunan laporan hasil perlombaan yang diawali dengan arahan dari supervisor, yaitu Bapak Raymond Sunardi Oetama dan dosen pembimbing, yaitu Bapak Samuel Ady Sanjaya agar mahasiswa dapat mulai mengerjakan laporan hasil lomba dengan lebih terarah dan terstruktur. Bimbingan atau *counseling* ini dilaksanakan di antara beberapa hari weekdays dalam rangka agar dosen dan *supervisor* mengetahui *progress* mahasiswa dalam penyusunan laporan hasil lomba.

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

Tabel 3.0.1 *Gantt Chart* Proyek Independen

Rentang kegiatan	Pekerjaan yang dilakukan	Agustus		September					Oktober				November			
		4	5	1	2	3	4	5	1	2	3	4	1	2	3	4
Pra-Turnamen	Mengumpulkan informasi mengenai kompetisi yang dapat diregistrasi untuk program <i>Project Independent Acceleration</i>															
	Mengumpulkan informasi mengenai kompetisi Turnamen Sains Data Nasional (TSDN)															
	<i>Literatur review</i> secara komprehensif mengenai data sains															
Turnamen	Menentukan topik dan data yang akan digunakan															
	Pre-Processing data															
	Mendalami algoritma K-Means															
	Pengerjaan <i>Social Network Analysis (SNA): Sheet Edges</i> dan visualisasi															
Pasca Turnamen	Penulisan proposal dan pengumpulan dokumen lomba															

3.2 Fase akhir yang akan dicapai

Pengerjaan *Social Network Analysis* (SNA) dengan metrik sentralitas *degree*, *betweenness*, *closeness*, dan *eigenvector centrality* menggunakan Google Colab menghasilkan sejumlah metrik sentralitas untuk setiap akun dalam jaringan sosial. Metrik *degree* memberikan informasi tentang sejauh mana suatu akun terhubung dengan akun lain dalam jaringan. Metrik *betweenness* mengukur seberapa sering suatu akun berada pada jalur terpendek antara akun lain, menunjukkan potensi peran "jembatan" atau penghubung dalam aliran informasi. *Closeness centrality* menunjukkan seberapa dekat suatu akun dengan akun lainnya, sementara *eigenvector centrality* memberikan skor berdasarkan koneksi akun dengan akun-akun berpengaruh. Hasil metrik-metrik ini dapat digunakan untuk mengidentifikasi akun-akun yang mungkin mencurigakan dalam jaringan.

Selanjutnya, hasil SNA dapat divisualisasikan menggunakan Gephi, alat analisis jaringan yang kuat. Gephi memungkinkan penyajian grafis yang intuitif dan dinamis dari jaringan sosial, dengan memanfaatkan warna, ukuran node, dan lebar *edge* untuk mencerminkan metrik sentralitas. Dengan visualisasi ini, peneliti dapat dengan mudah mengidentifikasi cluster, hubungan, dan pola-pola dalam jaringan. Dengan menerapkan filter dan menggambarkan akun-akun dengan metrik sentralitas tertentu, seperti *degree* yang tinggi atau posisi *betweenness* yang mencurigakan, Gephi mempermudah deteksi akun-akun sosial media yang patut dicurigai. Kombinasi antara analisis metrik sentralitas menggunakan Google Colab dan visualisasi dengan Gephi dapat menjadi alat yang efektif dalam memahami dan mengidentifikasi perilaku mencurigakan dalam jaringan sosial media.

3.3 Koleksi data

Proses pengumpulan data merupakan tahapan kritis dalam sebuah penelitian atau analisis, dimulai dengan perencanaan yang cermat untuk mengidentifikasi tujuan dan jenis data yang diperlukan. Desain instrumen pengumpulan data, seperti kuesioner atau panduan wawancara, harus diperhatikan agar sesuai dengan pertanyaan penelitian dan dapat menghasilkan data yang valid. Pemilihan sampel yang representatif juga menjadi aspek penting, dengan penerapan metode pemilihan sampel yang tepat untuk memastikan representasi yang baik dari populasi. Selanjutnya, pelaksanaan pengumpulan data membutuhkan pengawasan yang teliti sesuai dengan rencana dan petunjuk yang telah ditetapkan. Validasi dan verifikasi data menjadi langkah kunci untuk memeriksa kualitas data agar akurat dan lengkap.

Setelah data terkumpul, tahap selanjutnya melibatkan pengolahan dan analisis data. Proses ini mencakup pemasukan data ke dalam sistem pengolahan data, diikuti oleh langkah-langkah awal seperti pembersihan dan transformasi data. Analisis data kemudian dilakukan dengan menggunakan alat atau metode analisis yang sesuai dengan tujuan penelitian atau analisis yang telah ditentukan. Hasil analisis tersebut kemudian diinterpretasikan dan disajikan dalam laporan yang komprehensif, memungkinkan peneliti atau analis untuk menyampaikan temuan dengan jelas kepada audiens yang dituju. Dengan perencanaan yang matang dan implementasi yang hati-hati, proses pengumpulan data menjadi landasan yang kuat bagi keberhasilan sebuah penelitian atau analisis. Populasi mencakup semua subjek penelitian atau objek yang menjadi fokus penelitian [64]. Dalam konteks penelitian ini, populasi merujuk pada keseluruhan data yang terkait dengan pengguna aktif platform X yang terlibat dalam interaksi yang berkaitan dengan topik Pilpres 2024. Seluruh data tersebut akan menjadi dasar untuk analisis dan pemahaman dalam penelitian ini. Sampel merujuk pada metode atau prosedur yang digunakan untuk memilih sejumlah item atau individu yang lebih kecil dari jumlah populasi yang telah ditetapkan, dengan maksud untuk menggunakan mereka sebagai sumber data dalam penelitian [65]. Dalam penelitian ini, metode pengambilan sampel yang

diterapkan adalah purposive sampling, di mana pemilihan sampel dilakukan berdasarkan karakteristik yang dianggap relevan dengan tujuan penelitian [66]. Sampel yang digunakan dalam penelitian ini merupakan data yang berasal dari platform X, dengan fokus khusus pada informasi akun yang terlibat dalam interaksi atau memiliki keterkaitan dengan portal berita yang membahas topik pilpres.

Data yang diperoleh didapat melalui kegiatan *scraping*. *Web scraping* dari Twitter adalah proses ekstraksi data atau informasi dari halaman-halaman web Twitter menggunakan teknik *scraping*. Adapun kegiatan *scraping* ini menggunakan tools NodeXL yang merupakan plugin dari Microsoft Excel. Adapun dalam *scraping*, batasannya adalah sebagai berikut:

1. *Tweet* yang dipakai dalam rangka melakukan analisis merupakan *tweet* dengan kata kunci “pilpres” yang berbahasa Indonesia dan diunggah oleh WNI.
2. Rentang waktu *Tweet* yang di-*scrap* adalah tanggal 1 Januari – 21 Mei 2023 berjumlah 1.000.000 data, namun data yang digunakan untuk *Social Network Analysis* adalah data pada tanggal 1 Januari hingga 7 Januari 2023 yang berjumlah 17.132.
3. Analisis data *tweet* yang dilakukan mencakup interaksi percakapan melalui *tweet*, *retweet*, dan *mention* dari akun yang menyebarkan informasi terkait dengan proses pemilihan presiden.

3.3 Penyusunan desain teknis

1. Mengumpulkan informasi mengenai kompetisi nasional yang dapat diregistrasi sebagai program *Project Independent Acceleration*

Beberapa langkah komprehensif untuk mengumpulkan informasi tentang kompetisi nasional yang dapat didaftarkan sebagai bagian dari program *Project Independent Acceleration* di antaranya sebagai berikut:

- a. Menetapkan tujuan

Tujuan tersebut untuk dilihat kesesuaiannya untuk mengukur tingkat keahlian, mencari pengakuan nasional, atau mendukung pengembangan proyek independen. Penetapan tujuan akan membimbing proses pencarian informasi.

a. Identifikasi sumber informasi

Kegiatan ini merupakan kegiatan dalam menentukan sumber informasi yang dapat diandalkan. Hal ini mencakup termasuk situs web resmi pemerintah, portal kompetisi, atau lembaga pendidikan dan riset yang mendukung proyek independen.

b. Pemeriksaan situs web resmi

Kegiatan ini merupakan kegiatan untuk memeriksa situs web resmi pemerintah atau lembaga yang mengelola kompetisi nasional. Situs ini biasanya menyediakan informasi terkait syarat dan ketentuan, kriteria penilaian, tanggal pendaftaran, dan hadiah yang dapat dimenangkan.

c. Menghubungi penyelenggara kompetisi

Beberapa informasi yang diperoleh dari situs web tidak mencukupi, maka dari itu kegiatan ini dilakukan untuk menghubungi penyelenggara kompetisi melalui email, telepon, atau media sosial. Pihak-pihak tersebut memberikan panduan lebih lanjut dan menjawab pertanyaan yang dimiliki.

d. Riset pengalaman peserta sebelumnya

Kegiatan ini dilakukan dengan mencari ulasan atau pengalaman peserta sebelumnya dalam kompetisi tersebut. Hal ini dapat memberikan wawasan tentang tingkat kesulitan, nilai tambah, dan manfaat yang dapat diperoleh melalui partisipasi.

e. Eksplorasi komunitas *online*

Kegiatan ini dilakukan dengan bergabung di komunitas *online* terkait bidang proyek independen atau kompetisi. Forum, grup media sosial, atau platform diskusi dapat menjadi sumber informasi berharga, dan dapat berbagi pengalaman dengan orang lain yang memiliki minat serupa.

f. Memerhatikan syarat dan ketentuan

Kegiatan ini dilakukan dengan membaca dan memahami dengan seksama syarat dan ketentuan kompetisi. Hal ini mencakup persyaratan pendaftaran, kriteria penilaian, batas waktu, dan segala hal yang perlu dipersiapkan.

g. Verifikasi kredibilitas

Kegiatan ini dilakukan agar kompetisi tersebut memiliki kredibilitas yang baik dan diakui di tingkat nasional. Informasi tentang sponsor, juri, atau dukungan dari lembaga terkemuka dapat memberikan indikasi tentang kredibilitasnya.

h. Mempersiapkan dokumen pendukung

Kegiatan ini merupakan kegiatan persiapan semua dokumen yang diperlukan untuk pendaftaran yang termasuk proposal proyek, CV, referensi, atau dokumen lain yang mungkin diminta oleh penyelenggara.

i. Mencatat informasi penting

Kegiatan ini adalah menyimpan semua informasi yang ditemukan dalam format yang mudah diakses. Ini mencakup tanggal penting, kontak penyelenggara, dan persyaratan kunci lainnya.

Dengan mengikuti langkah-langkah ini, dapat dikumpulkan informasi secara komprehensif tentang kompetisi nasional yang dapat diikuti sebagai bagian dari program *Project Independent Acceleration*.

2. Mengumpulkan informasi mengenai kompetisi Turnamen Sains Data Nasional (TSDN)

a. **Situs Web Resmi TSDN**

Dilakukan dengan memeriksa situs web resmi TSDN. Sebagian besar informasi resmi terkait dengan aturan, jadwal, dan pendaftaran biasanya diposting di situs web resmi acara. Di samping itu juga memeriksa tautan yang merujuk ke informasi terkait tahun 2021, 2022, atau 2023.

b. **Dokumentasi resmi**

Dilakukan dengan memeriksa dokumentasi resmi yang mungkin telah dipublikasikan, seperti brosur, panduan peserta, atau dokumen lainnya yang memberikan pandangan komprehensif tentang acara tersebut.

c. **Media sosial**

Dilakukan dengan menjelajahi platform media sosial resmi TSDN, seperti akun Twitter, Facebook, atau Instagram. Pada platform ini, didapatkan pembaruan terkini, informasi pendaftaran, dan serangkaian posting yang memberikan wawasan lebih lanjut tentang turnamen.

d. **Kontak dengan penyelenggara**

Jika informasi yang dibutuhkan tidak tersedia secara *online*, maka menghubungi penyelenggara acara. Pihak tersebut dapat memberikan informasi langsung atau merujuk ke sumber daya yang relevan.

e. **Bergabung dengan komunitas sains data.**

Hal ini dilakukan dengan terhubung dengan komunitas *online* atau forum yang fokus pada sains data. Peserta sebelumnya atau mereka yang memiliki minat serupa dapat memberikan wawasan berharga tentang pengalaman mereka dan memberikan informasi praktis tentang kompetisi.

f. **Pencarian berita dan artikel**

Dilakukan dengan pencarian berita atau artikel *online* terkait TSDN. Media atau publikasi industri mungkin telah melaporkan tentang acara tersebut, memberikan pandangan eksternal yang dapat membantu memahami lebih baik turnamen tersebut.

g. **Pertanyaan pada peserta sebelumnya**

Jika ada forum atau platform di mana peserta sebelumnya berkomunikasi, ajukan pertanyaan kepada mereka. Pengalaman langsung dari peserta sebelumnya dapat memberikan wawasan yang berharga.

h. **Periksa hasil dan laporan tahun sebelumnya**

Dilakukan dengan menelusuri hasil dan laporan dari tahun-tahun sebelumnya. Ini dapat memberikan gambaran tentang jenis tantangan, tingkat kesulitan, dan hal-hal lain yang dapat diharapkan peserta.

Di samping itu juga memastikan untuk memeriksa informasi ini secara berkala, karena detail acara dan persyaratan dapat berubah dari tahun ke tahun. Dengan menggabungkan sumber informasi yang berbeda dapat terbentuk pemahaman yang lebih komprehensif tentang Turnamen Sains Data Nasional pada tahun-tahun yang diinginkan. Dari pengumpulan informasi yang sudah dilakukan, tabel 3.1 menunjukkan informasi ringkasan penyelenggaraan TSDN dari tahun 2021, 2022, hingga tahun 2023 (tahun ini).

A large, light blue watermark logo of Universitas Multimedia Nusantara (UMMN) is centered on the page. It features a stylized globe with a grid of squares and the acronym 'UMMN' in a bold, rounded font.

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

Tabel 3.0.2 Ringkasan Informasi TSDN dari Tahun 2021 hingga 2023

Tahun Penyelenggaraan	Tema	Tanggal dimulai	Tanggal berakhir	Sponsor	Support		
					Pemerintah	Asosiasi	Edukasi
2021	Beberapa topik <i>Use Case</i> : kesehatan masyarakat, sosial ekonomi, pelayanan publik, dan UMKM.	10 September 2021	28 Oktober 2021	Data Academy, Jakarta Open Data, Jabar Digital Service, Open Data Jabar	Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, Diskominfo DKI Jakarta, Diskominfo Jabar	Asosiasi Data Sains dan AI Indonesia, Asosiasi Ilmuwan Data Indonesia, ABOI (Asosiasi Big Data Indonesia), Id Big Data	Universitas Multimedia Nusantara (UMN), UII, UIN, Telkom University, UNNES, Universitas Mikroskil, UAI, dan lembaga edukasi lainnya
2022	Implementasi Sains Data dan AI untuk Bisnis	5 Oktober 2022	1 Desember 2022	Data Academy, Cybertrend, karier.mu, Safarindo, Good Choice, LSP, Paraogon, Lojin Organic Store	Kominfo, BNSP, BRIN, Jakarta Satu, Jakarta Open Data, Jabar Digital Service, Open Data Jabar	ABOI (Asosiasi Big Data Indonesia), Asosiasi Data Sains dan AI Indonesia, APTIKOM, Id Big Data, Asosiasi Ilmuwan Data Indonesia	UMN, Perbanas Institute, Universitas Inaba, UBM, Universitas Bakrie, SGU, dan instansi edukasi lainnya
2023	“Meningkatkan Keamanan Menggunakan	10 Oktober 2023	13 Desember 2023	LPS, Data Academy, Pyfa Health, uptodate studio	PANRB, DJP, Satu Data Jakarta, Jabar Digital Service, Open Data Jabar	Asosiasi Data Sains dan AI Indonesia, <i>data science</i> Indonesia,	Universitas Multimedia Nusantara, Universitas Trisakti,

Tahun Penyelenggaraan	Tema	Tanggal dimulai	Tanggal berakhir	Sponsor	Support		
					Pemerintah	Asosiasi	Edukasi
	Keilmuan Sains Data dan AI'			community, Stifin Family, Tanam Duit, Karier.mu, NF Computer		ABOI (Asosiasi Big Data Indonesia), Id Big Data	ESQ Business School, UIN, UII, Universitas Bakrie, SMK Telkom Malang, Telkom University, dan beberapa lembaga edukasi lainnya.

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

3. *Literatur review* secara komprehensif mengenai data sains

Kegiatan ini dilakukan dengan tinjauan pustaka yang menyeluruh dan mendalam terkait dengan berbagai aspek dan konsep dalam bidang data sains. Proses ini melibatkan pembacaan dan analisis literatur-literatur yang relevan, baik yang sudah ada maupun yang masih terkini, untuk memahami dan menggambarkan lanskap pengetahuan terkini seputar data sains. Literatur review ini mencakup pemahaman mendalam tentang metode-metode, teknik analisis, perkembangan terkini, tren, serta isu-isu kritis yang berkaitan dengan pengolahan, analisis, dan interpretasi data. Secara komprehensif berarti melibatkan berbagai sumber informasi seperti jurnal ilmiah, buku teks, makalah konferensi, dan sumber literatur akademis lainnya. Tujuan dari *literature review* ini adalah untuk membangun pemahaman yang mendalam tentang kerangka kerja konseptual, alat, teknik, dan isu-isu etis yang berkaitan dengan data sains. Dengan demikian, tinjauan pustaka yang komprehensif dapat memberikan dasar pengetahuan yang kuat bagi peneliti atau praktisi yang terlibat dalam bidang data sains, serta membantu mengidentifikasi area-area penelitian yang belum tersentuh atau memerlukan penelitian lebih lanjut.

Penelitian yang dilaksanakan oleh Arunadevi, J., & Sowjanya, A. M. pada tahun 2022 fokus pada pendeteksian berita palsu dengan menggunakan kombinasi algoritma klustering dan klasifikasi. Penelitian ini mengevaluasi headline dari berita-berita sebagai objek penelitian. Dalam perbandingan antara dua algoritma klustering, yakni K-Means dan K-Medoid, hasil penelitian menunjukkan bahwa performa K-Means lebih unggul dalam mengklusterisasi dibandingkan dengan K-Medoid [12].

Studi yang dilakukan oleh Samuel Ady Sanjaya dan Kridanto Surendro mengadopsi teknik PCA serta klustering K-means dengan jarak Mahalanobis untuk mengidentifikasi pesan spam dalam kumpulan data Twitter sebanyak 150 ribu. Ada kekurangan dalam klasifikasi spam yang hanya memisahkan data menjadi dua kelas, yakni spam dan bukan spam, sementara kelas lainnya tidak terdefiniskan. Model yang dihasilkan mengelompokkan cluster menjadi lima:

normal, akun berita, akun asing, tokoh masyarakat, dan spam. Melalui penggunaan kombinasi k-means dan PCA, dapat diidentifikasi cluster spam secara lebih spesifik dibandingkan dengan metode *supervised learning*. Sejumlah 6.230 akun teridentifikasi sebagai spam, sementara 3.374 akun termasuk dalam kategori pengguna aktif dan akun berita [24].

Penelitian berjudul "*Classification of Fake News on Facebook: A Novel Social Network with K-Means Clustering Approach for Against Principal Component Analysis Method for Better Accuracy*" melakukan klasifikasi berita palsu di Facebook dengan memanfaatkan algoritma pembelajaran mesin yang menawarkan tingkat akurasi yang lebih baik. Metode dan bahan dari penelitian ini melibatkan klasifikasi berita palsu yang diterapkan pada dataset, dengan tujuan mendeteksi makna sebenarnya dari konten tersebut. Dalam penelitian ini, dataset diberi label berupa judul, teks, subjek, dan tanggal. Data tersebut kemudian diaplikasikan pada algoritma pembelajaran mesin seperti K-Means (KM) Clustering, yang diambil sebagai kelompok-1, dan dibandingkan dengan algoritma *Principal Component Analysis* (PCA), yang diambil sebagai kelompok-2. Hasil dari penelitian ini menunjukkan bahwa jejaring sosial baru yang digunakan dalam eksperimen ini, dengan penerapan *clustering* K-Means dan algoritma PCA, menghasilkan prediksi algoritma terbaik dengan tingkat akurasi yang lebih tinggi [25].

Kesimpulan dari tiga penelitian yang telah diuraikan menunjukkan variasi pendekatan dalam menggunakan teknik data sains untuk mengatasi tantangan spesifik. Penelitian pertama oleh Arunadevi, J., & Sowjanya, A. M. tahun 2022 menggambarkan keunggulan algoritma K-Means dalam mengklusterisasi headline berita palsu dibandingkan dengan K-Medoid. Hal ini menunjukkan pentingnya pemilihan algoritma klustering yang tepat dalam mendeteksi dan mengelompokkan berita palsu. Studi kedua oleh Samuel Ady Sanjaya dan Kridanto Surendro mengadopsi PCA dan K-means clustering untuk mengidentifikasi spam di Twitter. Meskipun terdapat kekurangan dalam klasifikasi yang hanya membagi data menjadi dua kelas, model yang dihasilkan

mampu mengelompokkan *cluster* menjadi lima kategori, memperlihatkan potensi penggunaan kombinasi K-means dan PCA untuk identifikasi cluster yang lebih spesifik. Penelitian terakhir, yang berfokus pada klasifikasi berita palsu di Facebook, menunjukkan upaya untuk meningkatkan akurasi dengan menggunakan algoritma pembelajaran mesin. Penggunaan K-Means *Clustering Approach* dan *Principal Component Analysis Method* pada dataset berlabel judul, teks, subjek, dan tanggal menghasilkan prediksi algoritma terbaik dengan tingkat akurasi yang lebih tinggi. Secara keseluruhan, kesimpulan ini menekankan perlunya pemilihan metode dan algoritma yang tepat, sekaligus menunjukkan beragamnya pendekatan yang dapat digunakan dalam mendukung pengembangan teknik deteksi dan klasifikasi dalam konteks berita palsu dan spam di platform sosial media.

4. *Pre-processing data*

Pre-processing data dalam konteks deteksi akun mencurigakan (*suspicious account detection*) menggunakan algoritma *machine learning* adalah tahap awal yang kritis dalam pengembangan sistem tersebut. Proses ini melibatkan serangkaian langkah untuk membersihkan, mengorganisir, dan mempersiapkan data mentah agar dapat diolah dengan efektif oleh algoritma *machine learning*. Berikut adalah langkah-langkah komprehensif untuk *pre-processing data* pada topik *suspicious account detection*:

a. **Pemahaman data**

Dilakukan dengan identifikasi sumber data yang relevan untuk deteksi akun mencurigakan. Pahami struktur data, variabel-variabel yang ada, dan format data.

b. **Pembersihan data**

Dilakukan dengan identifikasi dan tangani nilai-nilai yang hilang atau tidak valid. Deteksi dan tangani *outlier* yang mungkin memengaruhi hasil deteksi mencurigakan.

c. **Normalisasi data**

Dilakukan dengan menormalisasikan data numerik untuk memastikan konsistensi skala. Terapkan teknik normalisasi seperti *Min-Max Scaling* atau *Z-score normalization*.

d. Ekstraksi fitur

Dilakukan dengan identifikasi fitur-fitur yang mungkin relevan untuk deteksi akun mencurigakan. Ekstraksi fitur-fitur seperti aktivitas login, frekuensi posting, atau pola interaksi.

e. Manajemen data teks

Jika data mencakup teks (misalnya, posting di media sosial), lakukan tokenisasi dan penghapusan *stop words*. Dilakukan dengan menerapkan teknik vektorisasi teks seperti TF-IDF atau *Word Embeddings*.

f. Penanganan ketidakseimbangan kelas

Periksa dan tangani ketidakseimbangan dalam distribusi kelas antara akun normal dan mencurigakan. Dilakukan dengan menerapkan teknik *oversampling* atau *undersampling* sesuai kebutuhan.

g. Manajemen waktu

Jika data mencakup *timestamp*, dilakukan ekstraksi fitur waktu dan identifikasi pola temporal yang mencurigakan.

h. Penghapusan fitur tak berguna

Dilakukan dengan identifikasi dan hapus fitur yang tidak memberikan informasi berharga atau menyebabkan *multicollinearity*.

i. Penanganan duplikasi data

Dilakukan dengan identifikasi dan tangani data duplikat agar tidak memengaruhi hasil analisis.

Dengan melakukan *pre-processing data*, didapatkan kualitas input yang meningkat yang akan digunakan algoritma *machine learning*, yang pada gilirannya dapat meningkatkan kinerja sistem deteksi akun mencurigakan. Proses ini dapat membantu mencegah masalah seperti *overfitting*, meningkatkan interpretabilitas model, dan mengoptimalkan hasil deteksi.

5. Menentukan topik dan data yang akan digunakan 3.

Topik yang diangkat dalam penelitian ini berfokus pada deteksi akun yang mencurigakan dengan tujuan mengatasi permasalahan berita hoaks dan dampak negatif lainnya, seperti *phishing*, *cyber bullying*, dan *cyber crime*, yang dapat ditimbulkan oleh akun-akun tersebut. Pendekatan yang diambil melibatkan pengumpulan data dari media sosial X dengan menggunakan teknik *scraping*, menggunakan alat bantu *NodeXL*, dengan kata kunci "pilpres" dan dalam periode waktu tertentu. Metode pengolahan data melibatkan penggunaan algoritma K-means untuk klusterisasi akun-akun tersebut. Sebagai upaya untuk mereduksi dimensi pada data yang kompleks, metode *Principal Component Analysis* (PCA) diterapkan. Dengan mengimplementasikan PCA, diharapkan proses klusterisasi menjadi lebih efisien dan akurat, memungkinkan identifikasi akun mencurigakan secara lebih tepat dan responsif terhadap konten hoaks serta ancaman siber lainnya yang dapat muncul di platform media sosial. Dengan demikian, penelitian ini mengeksplorasi integrasi teknologi dan metode analisis data untuk mengatasi permasalahan serius yang muncul dalam lingkungan digital saat ini.

6. Pengerjaan SNA dan visualisasinya

Dilakukan pengerjaan SNA menggunakan tools Google Collab dan visualisasi hasil SNA menggunakan Gephi

7. Penulisan proposal dan dokumen lomba

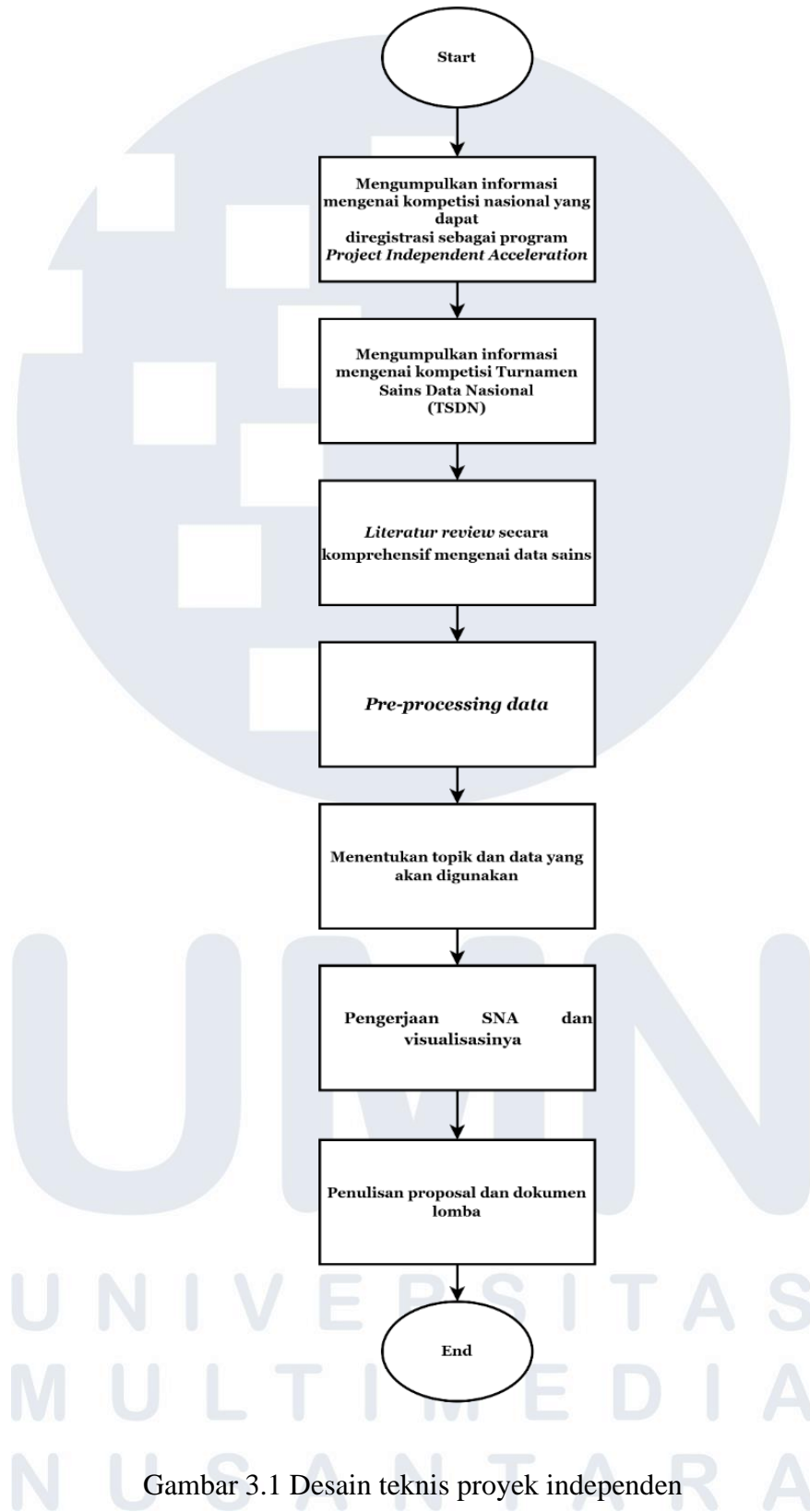
Proses penulisan proposal dan pengumpulan dokumen lomba merupakan tahapan krusial dalam berpartisipasi dalam kompetisi atau lomba. Penulisan proposal memerlukan kemampuan untuk merinci secara jelas tujuan, metodologi, dan dampak yang diharapkan dari proyek atau solusi yang diajukan. Proposal harus mampu meyakinkan para juri mengenai relevansi, inovasi, dan potensi implementasi dari ide atau konsep yang diajukan. Pengumpulan dokumen lomba, seperti laporan proyek, analisis data, atau presentasi,

membutuhkan ketelitian dan kesisteman agar informasi dapat disajikan dengan jelas dan meyakinkan. Ketersediaan dokumentasi yang lengkap dan akurat menjadi kunci kesuksesan dalam membuktikan nilai proyek atau solusi yang diusulkan. Oleh karena itu, peserta harus memastikan bahwa setiap dokumen yang diajukan mencerminkan kualitas dan dedikasi mereka dalam menghadapi tantangan lomba, serta memberikan gambaran komprehensif tentang kontribusi yang mereka bawa ke dalam kompetisi tersebut.

A large, light blue watermark logo of Universitas Multimedia Nusantara (UMMN) is centered on the page. It features a stylized globe with several white squares of varying sizes arranged in a grid-like pattern over it.

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA



Gambar 3.1 Desain teknis proyek independen