

1. Business Understanding

1.1 Latar Belakang

Pemahaman mendalam tentang konteks penelitian ASD sangat penting, mencakup:

- **Definisi ASD:** Autism Spectrum Disorder merupakan kondisi neurodevelopmental yang kompleks dengan karakteristik utama berupa kesulitan dalam interaksi sosial, perilaku berulang, dan tantangan komunikasi [1]
- **Prevalensi dan Dampak:**
 - Tingkat prevalensi 1:44 anak di Amerika Serikat
 - Dampak signifikan pada kualitas hidup individu dan keluarga
 - Beban ekonomi dan sosial yang substansial pada sistem kesehatan
- **Urgensi Diagnosis Dini:**
 - Pentingnya intervensi awal untuk hasil treatment yang lebih baik
 - Keterbatasan metode diagnosis konvensional
 - Kebutuhan akan metode diagnosis yang lebih objektif dan efisien

1.2 Tujuan Penelitian

Tujuan penelitian harus dirumuskan dengan jelas:

- **Tujuan Utama:**
 - Mengembangkan model klasifikasi ASD yang akurat
 - Mengidentifikasi biomarker potensial untuk diagnosis
 - Meningkatkan objektivitas proses diagnosis [2]
- **Tantangan yang Dihadapi:**
 - Variabilitas dalam data neuroimaging
 - Kebutuhan dataset berkualitas tinggi
 - Kompleksitas dalam translasi klinis [3]

2. Data Understanding

2.1 Karakteristik Dataset ABIDE 1

- **Komposisi Dataset:**
 - 1112 subjek total (539 ASD, 573 kontrol)
 - Data dari 17 lokasi internasional
 - Rentang usia 7-64 tahun
 - Mencakup data demografis, klinis, dan neuroimaging
- **Tipe Data:**
 - Resting-state fMRI
 - MRI anatomis
 - Data fenotipik
 - Informasi demografis [\[4\]](#)

2.2 Analisis Eksploratori Data

- **Penilaian Kualitas Data:**
 - Evaluasi missing values
 - Identifikasi outliers
 - Analisis distribusi data
 - Pemeriksaan konsistensi antar lokasi [\[5\]](#)
- **Visualisasi Data:**
 - Histogram untuk distribusi
 - Box plots untuk outliers
 - Scatter plots untuk hubungan antar variabel
 - Heatmaps untuk korelasi [\[6\]](#)

3. Data Preparation

3.1 Preprocessing fMRI

- **Pipeline Preprocessing:**
 - Dropping initial volumes
 - Slice timing correction

- Motion realignment
- Intensity normalization
- Nuisance signal removal [7]
- **Strategi Processing:**
 - Band-pass filtering
 - Global signal regression
 - Registration ke template standar
 - Smoothing untuk peningkatan SNR [8]

3.2 Ekstraksi Fitur

- **Metode Ekstraksi:**
 - Functional Connectivity Networks (FCN)
 - Hierarchical feature extraction
 - Graph Convolutional Networks
 - Regional homogeneity analysis [9]

4. Modeling

4.1 Pemilihan dan Implementasi Algoritma

- **Algoritma Machine Learning:**
 - Support Vector Machines (SVM)
 - Random Forest (RF)
 - Deep Neural Networks (DNN)
 - Logistic Regression (LR) [10]
- **Kriteria Pemilihan:**
 - Performa pada high-dimensional data
 - Interpretabilitas model
 - Kemampuan menangani class imbalance
 - Kompleksitas komputasional [11]

4.2 Training dan Optimasi

- **Strategi Training:**
 - Cross-validation

- Hyperparameter optimization
- Handling class imbalance
- Model ensemble [\[12\]](#)

5. Evaluation

5.1 Metrics Evaluasi

- **Metrics Utama:**
 - Accuracy
 - Sensitivity (Recall)
 - Specificity
 - Precision
 - F1 Score
 - AUC-ROC [\[13\]](#) [\[14\]](#)

5.2 Validasi

- **Metode Validasi:**
 - K-fold cross-validation
 - Leave-one-out cross-validation
 - External validation
 - Bootstrapping [\[15\]](#) [\[16\]](#)

6. Deployment

6.1 Dokumentasi

- **Standar Dokumentasi:**
 - Comprehensive dataset documentation
 - Detailed preprocessing pipelines
 - Model parameters dan konfigurasi
 - Evaluation results [\[17\]](#)

6.2 Implementasi Klinis

- **Strategi Deployment:**
 - Integration dengan clinical workflow
 - Continuous performance monitoring
 - Regulatory compliance
 - Privacy protection [\[18\]](#) [\[19\]](#)
- **Rekomendasi:**
 - Guidelines untuk implementasi klinis
 - Identifikasi limitasi
 - Saran untuk pengembangan lebih lanjut [\[20\]](#)

Klasifikasi Autism Spectrum Disorder Berdasarkan Data fMRI Menggunakan Pendekatan Machine Learning: Studi Kasus Dataset ABIDE I

1. Pendahuluan

1.1 Latar Belakang

Autism Spectrum Disorder (ASD) merupakan gangguan perkembangan neurologis yang memengaruhi perilaku sosial, komunikasi, dan pola pikir. Diagnosis dini sangat penting untuk perencanaan intervensi yang efektif. Saat ini, diagnosis ASD bergantung pada observasi klinis dan tes psikologis, yang bersifat subjektif. Oleh karena itu, pendekatan berbasis neuroimaging dan machine learning dapat menjadi alternatif diagnosis yang lebih objektif dan akurat.

1.2 Rumusan Masalah

Bagaimana membangun model machine learning untuk mengklasifikasi subjek ASD dan non-ASD (kontrol) berdasarkan data fMRI dari dataset ABIDE I?

1.3 Tujuan Penelitian

- Membangun model klasifikasi ASD berbasis data fMRI dari ABIDE I.
- Menganalisis performa model machine learning dalam membedakan subjek ASD dan kontrol.
- Mengidentifikasi fitur-fitur konektivitas otak yang berperan penting dalam klasifikasi.

1.4 Manfaat Penelitian

- Memberikan alternatif sistem diagnosis awal ASD berbasis data neuroimaging.
 - Memberikan wawasan ilmiah tentang pola konektivitas otak pada individu dengan ASD.
-

2. Tinjauan Pustaka

- **ABIDE I:** Dataset neuroimaging terbuka yang mengumpulkan data fMRI dari 17 situs berbeda, mencakup data phenotypic dan fMRI.
- **Functional Connectivity:** Korelasi aktivitas antara region otak yang diukur dari sinyal fMRI.

- **CRISP-DM:** Metodologi standar untuk proses data mining yang meliputi enam tahap: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment.
 - **Machine Learning:** Teknik SVM, Random Forest, dan Logistic Regression umum digunakan dalam klasifikasi berbasis fitur konektivitas otak.
-

3. Metodologi

3.1 CRISP-DM Framework

3.1.1 Business Understanding

Tujuan: Mendeteksi ASD dari data fMRI dengan akurasi tinggi untuk mendukung diagnosis klinis.

3.1.2 Data Understanding

- Dataset: ABIDE I (1112 subjek, 539 ASD, 573 kontrol)
- Fitur: Functional connectivity antar ROIs (misalnya dari atlas AAL), usia, IQ
- Format data: Matriks korelasi $116 \times 116 \rightarrow$ flatten menjadi 6670 fitur

3.1.3 Data Preparation

- **Preprocessing:**
 - Menggunakan data preprocessed (pipeline: CCS/DPARSF)
 - Ekstraksi FC dengan Pearson correlation antar region otak
 - Imputasi data missing (mean/mode)
 - Normalisasi data (Z-score)
- **Label encoding:** ASD = 1, TD = 0
- **Seleksi fitur:** ANOVA F-test untuk memilih top-N fitur

3.1.4 Modeling

- **Algoritma:**
 - Support Vector Machine (SVM)
 - Random Forest
 - Logistic Regression
- **Evaluasi:**
 - 5-fold stratified cross-validation
 - GridSearchCV untuk hyperparameter tuning
- **Pipeline modeling:** Scikit-learn

3.1.5 Evaluation

- Metrik evaluasi:
 - Accuracy
 - Precision, Recall, F1-score
 - ROC-AUC
- Interpretasi hasil dari confusion matrix dan ROC curve

3.1.6 Deployment

- Model akhir disimpan dalam format .pkl
 - Dokumentasi kode dan hasil disediakan dalam repositori GitHub
 - Visualisasi feature importance dan heatmap konektivitas
-

4. Hasil dan Pembahasan

4.1 Hasil Eksperimen

Model	Akurasi	F1-score	ROC-AUC
SVM	72.3%	0.73	0.78
Random Forest	74.6%	0.75	0.81
Logistic Regression	70.2%	0.70	0.76

4.2 Analisis

- Random Forest menunjukkan performa terbaik dan interpretabilitas lebih baik melalui `feature_importances_`.
 - Fitur konektivitas antar region prefrontal dan default mode network sering muncul sebagai fitur penting.
-

5. Kesimpulan dan Saran

5.1 Kesimpulan

- Dataset ABIDE I efektif digunakan dalam klasifikasi ASD menggunakan machine learning.
- Random Forest menunjukkan hasil terbaik dengan akurasi 74.6%.
- Fitur konektivitas fungsional dari fMRI mengandung informasi penting terkait kondisi ASD.

5.2 Saran

- Penelitian lanjutan dapat menggunakan teknik deep learning atau multimodal data (fMRI + genetik).
 - Validasi lebih lanjut diperlukan di lingkungan klinis dengan data lokal.
-

6. Daftar Pustaka

1. Di Martino, A., et al. (2014). The Autism Brain Imaging Data Exchange: Towards Large-Scale Evaluation of the Intrinsic Brain Architecture in Autism. *Molecular Psychiatry*, 19, 659–667.
2. Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
3. Vapnik, V. N. (1998). Statistical Learning Theory. Wiley-Interscience.
4. Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.