

# IDENTIFIKASI ANEMIA BERDASARKAN COMPLETE BLOOD COUNT MENGUNAKAN ALGORITMA CATEGORICAL BOOSTING (CATBOOST)

Riset Informatika

Vika Rafi Ana | 22081010009

# LATAR BELAKANG & TUJUAN PENELITIAN

- Anemia merupakan masalah kesehatan global dengan prevalensi tinggi, termasuk di Indonesia
- Diagnosis anemia dilakukan melalui pemeriksaan Complete Blood Count (CBC), namun interpretasinya kompleks karena banyak parameter yang saling berinteraksi
- Pendekatan machine learning dapat membantu meningkatkan akurasi dan objektivitas klasifikasi anemia
- Tujuan penelitian: mengimplementasikan algoritma CatBoost untuk mengklasifikasikan jenis anemia berdasarkan parameter CBC dan mengevaluasi kinerjanya

# DATASET PENELITIAN

Dataset berasal dari Kaggle dengan judul :

*Anemia Types Classification*

Dataset awal : 1.281 data

Data setelah praprocessing : 1.232

15 Fitur dan 9 Kelas Target Diagnosis

```
Total data Train: (985, 14)
Total data Test: (247, 14)
Distribusi Y_test:
Diagnosis
Healthy 65
Normocytic hypochromic anemia 54
Normocytic normochromic anemia 51
Iron deficiency anemia 37
Thrombocytopenia 15
Other microcytic anemia 11
Leukemia 9
Macrocytic anemia 3
Leukemia with thrombocytopenia 2
```

```
... Mounted at /content/drive
Total awal data dan Kolom: (1281, 15)
Distribusi awal per kelas:
Diagnosis
Healthy 336
Normocytic hypochromic anemia 279
Normocytic normochromic anemia 269
Iron deficiency anemia 189
Thrombocytopenia 73
Other microcytic anemia 59
Leukemia 47
Macrocytic anemia 18
Leukemia with thrombocytopenia 11
Name: count, dtype: int64
```

```
Total data dan Kolom (Setelah Duplikasi Dihapus): (1232, 15)
Distribusi per kelas (Setelah Duplikasi Dihapus):
Diagnosis
Healthy 323
Normocytic hypochromic anemia 271
Normocytic normochromic anemia 255
Iron deficiency anemia 184
Thrombocytopenia 72
Other microcytic anemia 56
Leukemia 44
Macrocytic anemia 16
Leukemia with thrombocytopenia 11
Name: count, dtype: int64
```

# METODOLOGI PENELITIAN

Pra-pemrosesan:

- Penghapusan duplikasi data
- Pemisahan fitur(X) dan target(Y)
- Penanganan ketidakseimbangan kelas menggunakan class weighting
- Split data 80% latih dan 20% uji

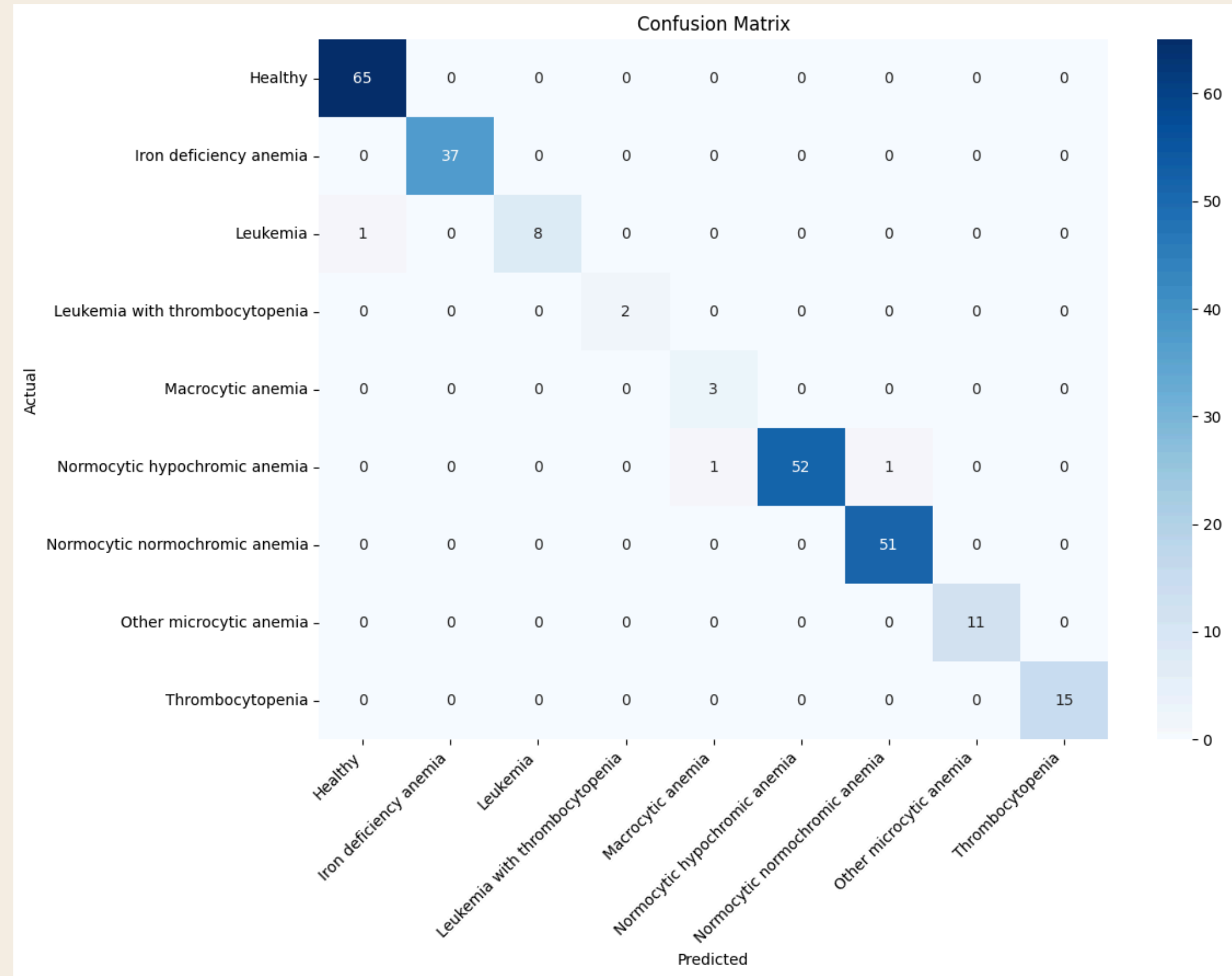
Pembangunan Model CatBoost

- Model: CatBoost Classifier
- Proses pelatihan menggunakan mekanisme gradient boosting dengan struktur oblivious decision tree
- Digunakan teknik ordered boosting untuk mengurangi risiko overfitting dan kebocoran data
- Evaluasi :
  - Akurasi, precision, recall, F1-score, dan confusion matrix

Akurasi: 0.9879

Classification Report:

	precision	recall	f1-score	support
Healthy	0.98	1.00	0.99	65
Iron deficiency anemia	1.00	1.00	1.00	37
Leukemia	1.00	0.89	0.94	9
Leukemia with thrombocytopenia	1.00	1.00	1.00	2
Macrocytic anemia	0.75	1.00	0.86	3
Normocytic hypochromic anemia	1.00	0.96	0.98	54
Normocytic normochromic anemia	0.98	1.00	0.99	51
Other microcytic anemia	1.00	1.00	1.00	11
Thrombocytopenia	1.00	1.00	1.00	15
accuracy			0.99	247
macro avg	0.97	0.98	0.97	247
weighted avg	0.99	0.99	0.99	247



# KESIMPULAN

Algoritma CatBoost terbukti efektif dalam mengklasifikasikan anemia berdasarkan parameter Complete Blood Count (CBC). Dengan pra-pemrosesan data dan penanganan ketidakseimbangan kelas, model mencapai akurasi 98,79% serta nilai precision, recall, dan F1-score yang tinggi pada hampir seluruh kelas. Hasil ini menunjukkan bahwa CatBoost berpotensi diterapkan sebagai sistem pendukung keputusan klinis untuk diagnosis anemia yang cepat dan akurat.