



# AST

PREDIKSI KUALITAS SUSU  
MENGUNAKAN METODE  
HYBRID ENSEMBLE DENGAN  
DATASET TAK SEIMBANG

KELOMPOK 6 - 2 D4 SDT

**Dosen Pengampu : Ronny Susetyoko S.Si., M.Si**

## NAMA ANGGOTA

- Puput Ayu Setiawati (3322600004)
- Bayu Kurniawan (3322600019)
- Abdul Muffid (3322600021)

# LATAR BELAKANG

Pada era industri yang berkembang pesat saat ini, pencapaian Sustainable Development Goals (SDGs) telah menjadi fokus utama dalam upaya meningkatkan keberlanjutan dan inovasi dalam berbagai sektor. Salah satu sektor yang memiliki peran penting dalam mencapai SDGs adalah industri susu. Industri susu berperan dalam memastikan kebutuhan gizi masyarakat terpenuhi dan meningkatkan perekonomian di berbagai negara. Namun, untuk mencapai tujuan inovasi industri dan memenuhi standar keamanan pangan yang ketat, perlu adanya sistem prediksi yang handal untuk mengukur dan mengendalikan kualitas susu secara efektif.

Prediksi kualitas susu menjadi elemen kunci dalam menjaga standar keamanan pangan, meningkatkan efisiensi produksi, dan mengurangi limbah. Inovasi dalam sektor ini menjadi sangat penting untuk mencapai tujuan SDGs yang berfokus pada pertumbuhan industri yang berkelanjutan. Dengan adanya perkembangan teknologi seperti machine learning dan analisis data, kita sekarang memiliki kemampuan untuk mengembangkan model prediksi yang lebih akurat untuk mengukur kualitas susu. Dalam konteks ini, penelitian tentang prediksi kualitas susu yang berbasis teknologi telah menjadi topik utama dalam upaya mendukung SDGs di sektor inovasi industri.

Penggunaan metode prediksi yang canggih, seperti Hybrid Ensemble yang menggabungkan berbagai algoritma seperti Logistik Ordinal, Random Forest, dan K-Nearest Neighbors, menjadi salah satu pendekatan yang menjanjikan dalam menghadapi tantangan prediksi kualitas susu. Dengan demikian, penelitian ini bertujuan untuk menyediakan kontribusi yang berharga dalam menggerakkan inovasi industri susu, menciptakan produk yang lebih berkualitas, dan mendukung pencapaian SDGs dalam sektor inovasi industri yang berkelanjutan.

## TUJUAN

- Mengidentifikasi Metode Ensemble Terbaik

- Mengukur Kinerja Metode Ensemble

- Mengidentifikasi Perbedaan Antar Model

- Memberikan Informasi Relevan



# DATASET

Dataset yang digunakan merupakan dataset yang membahas tentang produk susu. Untuk nama datasetnya adalah milk.csv yang sudah di processing menjadi milk\_fix.csv dan dataset ini diambil dari data kaggle. Dalam dataset tersebut memiliki 8 kolom atau 8 variabel, meliputi pH, Temperature, Taste, Odor, Fat, Turbidity, Colour, dan Grade.

Link sumber : <https://www.kaggle.com/datasets/cpluzshrijayan/milkquality>

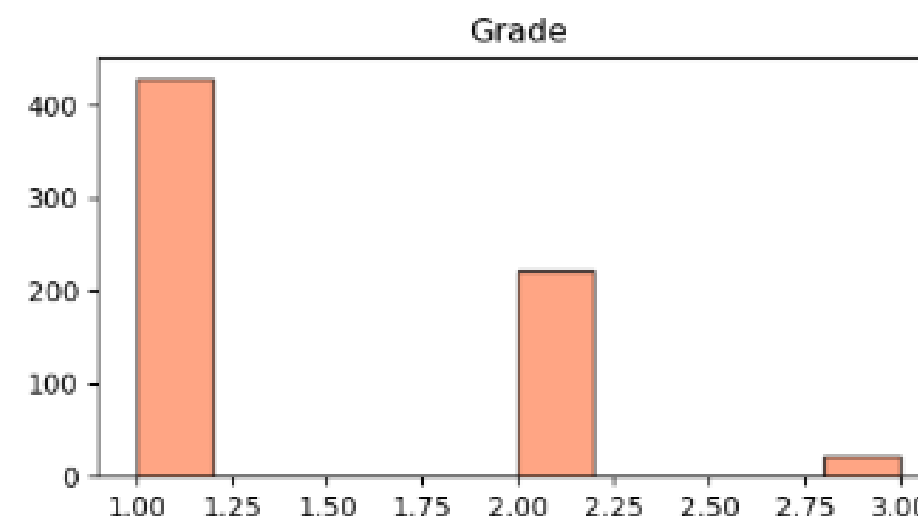
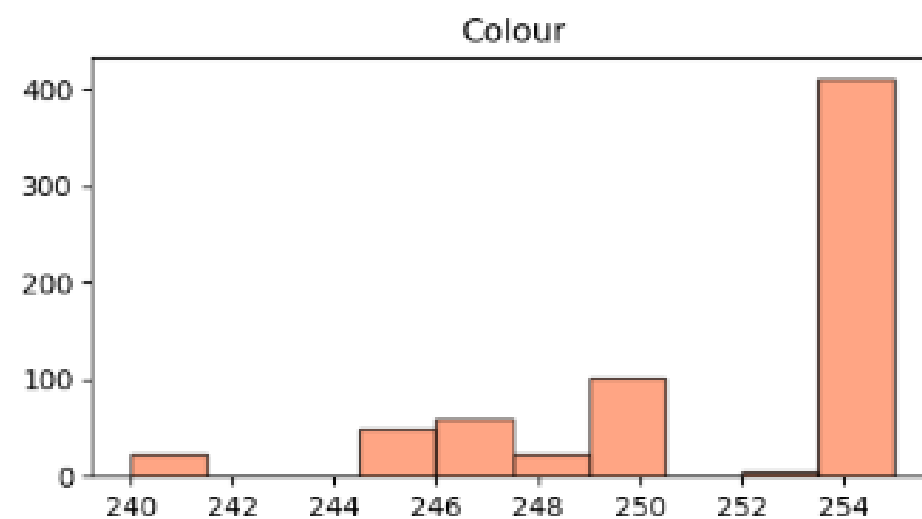
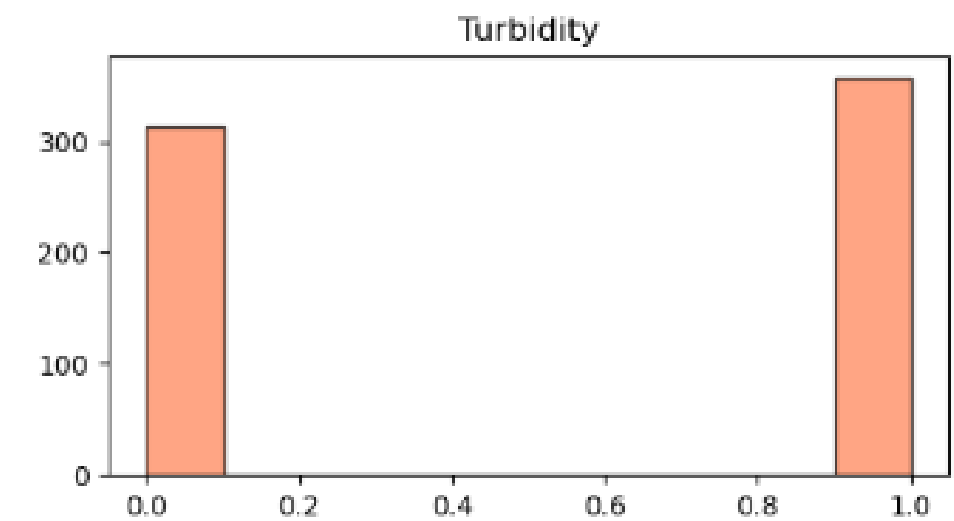
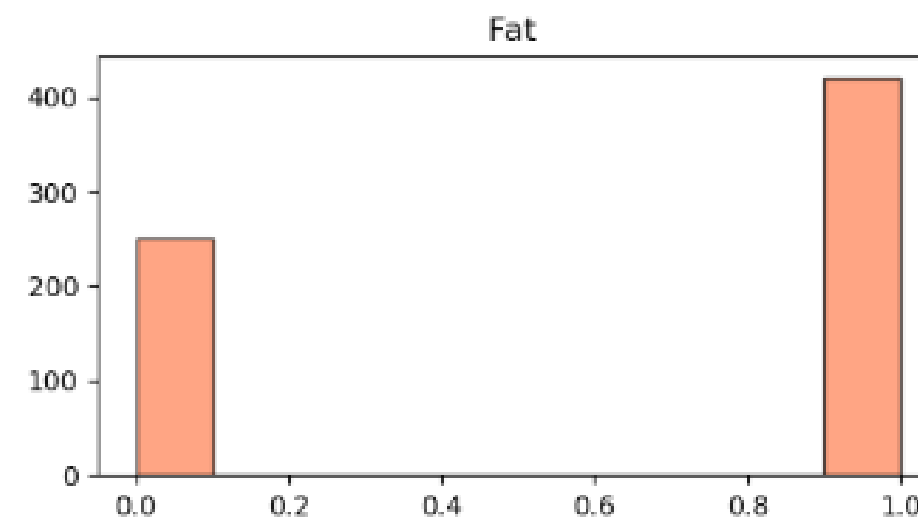
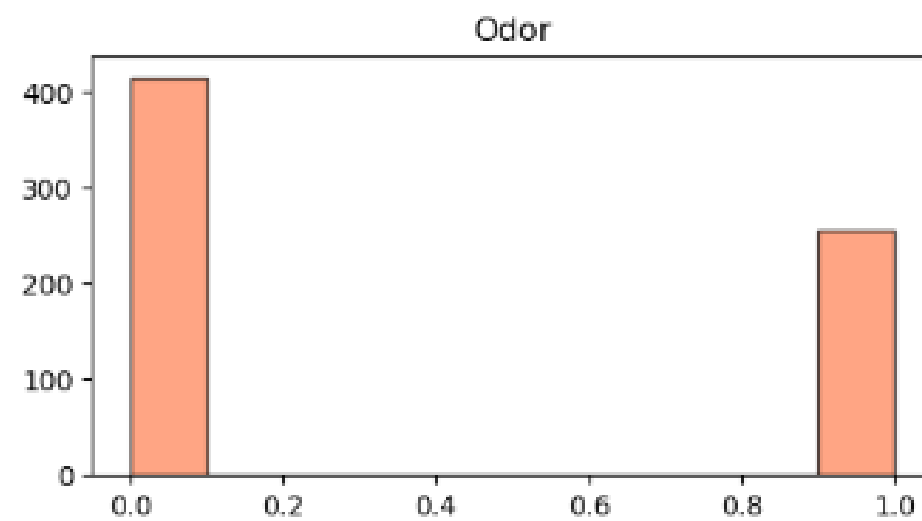
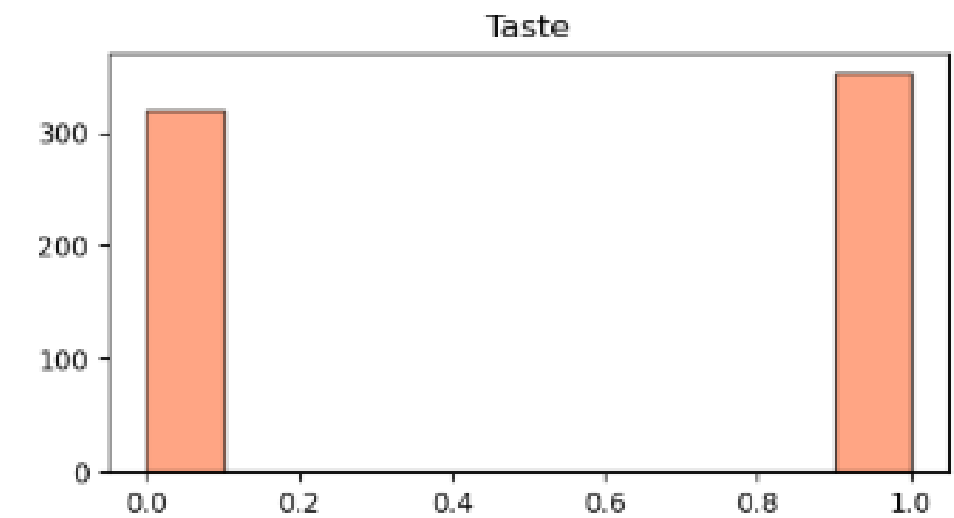
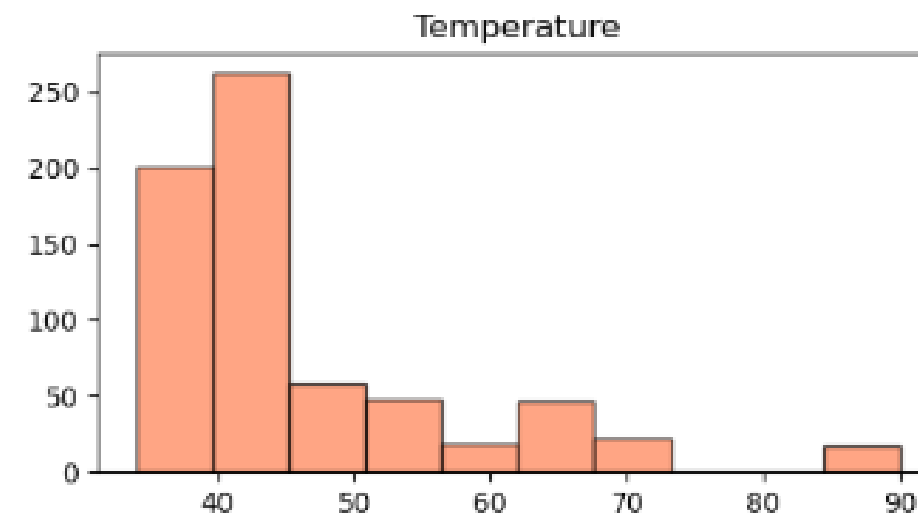
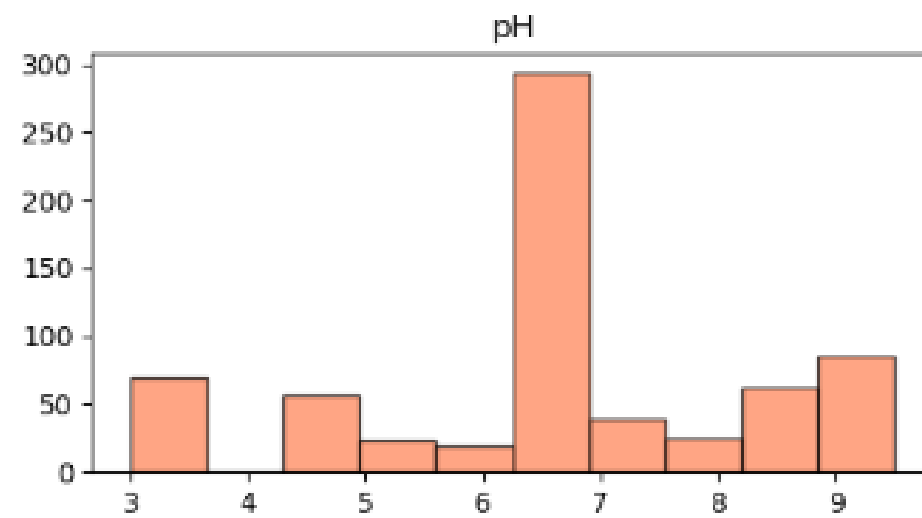
Link dataset : <https://intip.in/DataMilk>

# DATASET

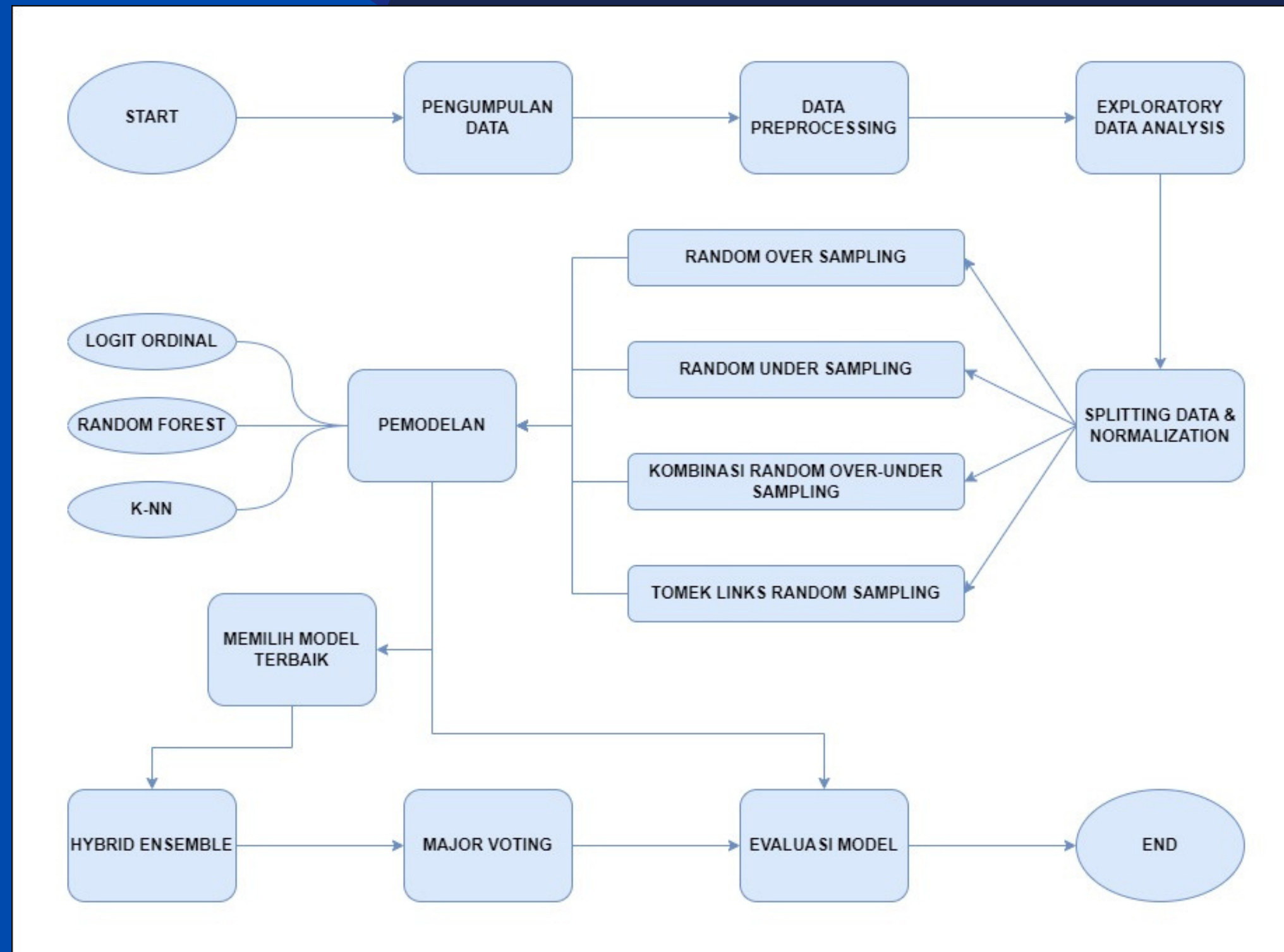
	pH	Temprature	Taste	Odor	Fat	Turbidity	Colour	Grade
0	8.5	70	1	1	1	1	246	1
1	9.5	34	1	1	0	1	255	1
2	5.5	45	1	0	1	1	250	1
3	4.5	60	0	1	1	1	250	1
4	8.1	66	1	0	1	1	255	1
...	...	...	...	...	...	...	...	...
667	6.8	36	0	1	1	0	253	3
668	6.8	45	0	1	1	1	255	3
669	6.6	35	0	1	1	1	255	3
670	6.6	35	0	1	1	1	255	3
671	6.6	45	0	1	1	1	250	3

672 rows × 8 columns

# EXPLORATORY DATA ANALYSIS - HISTOGRAM

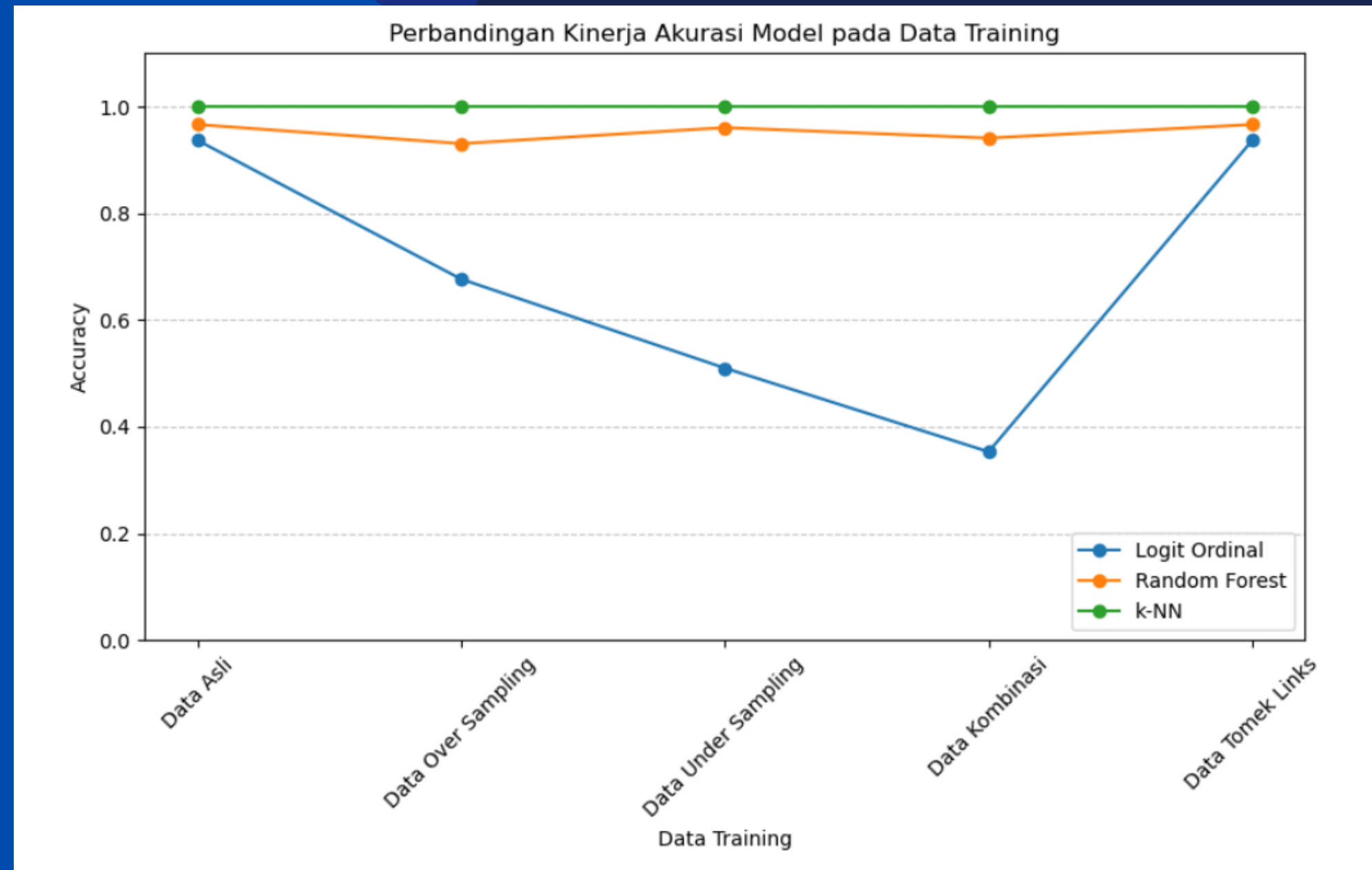


# METODOLOGI



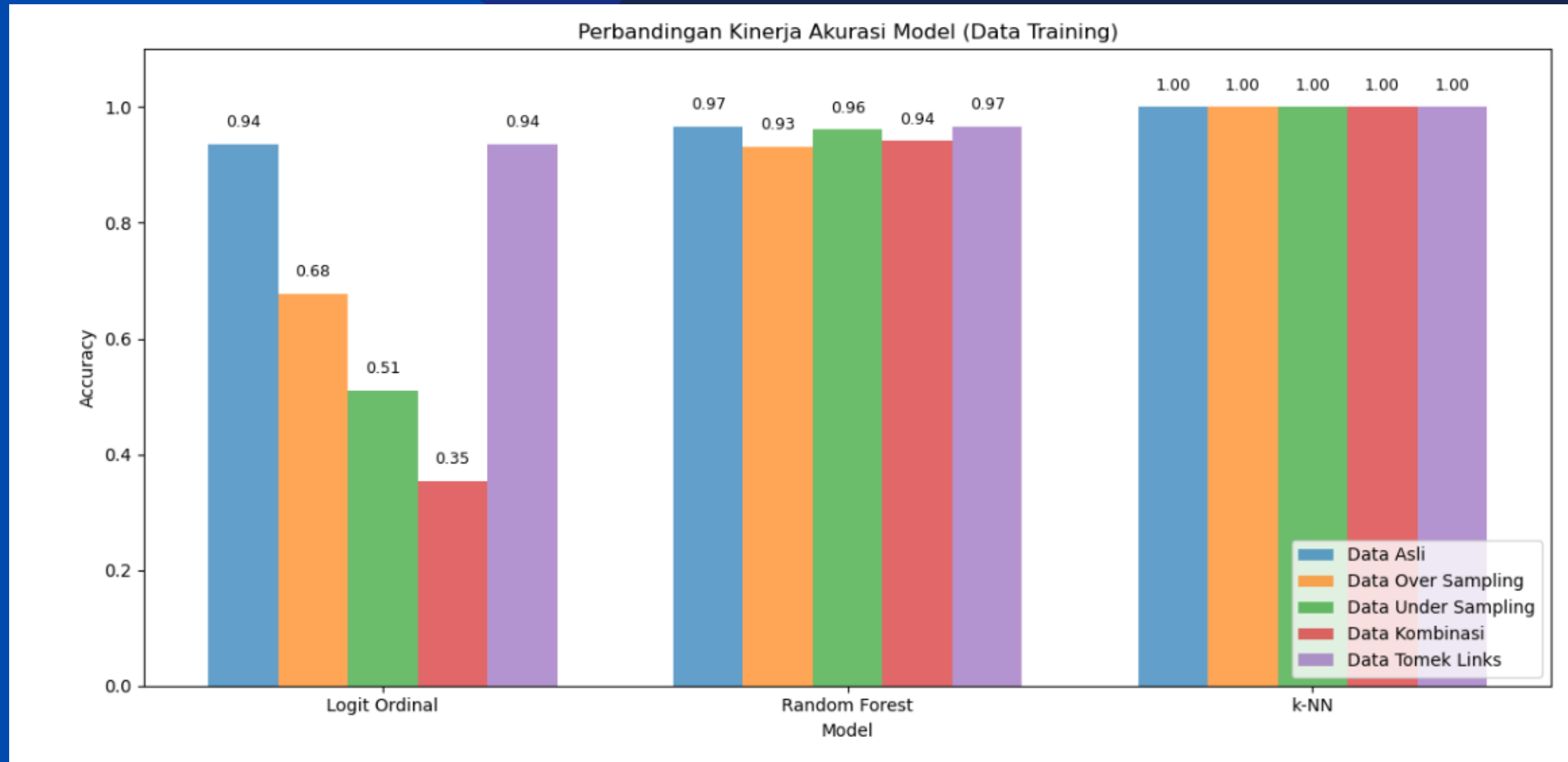


# AKURASI DATA TRAINING



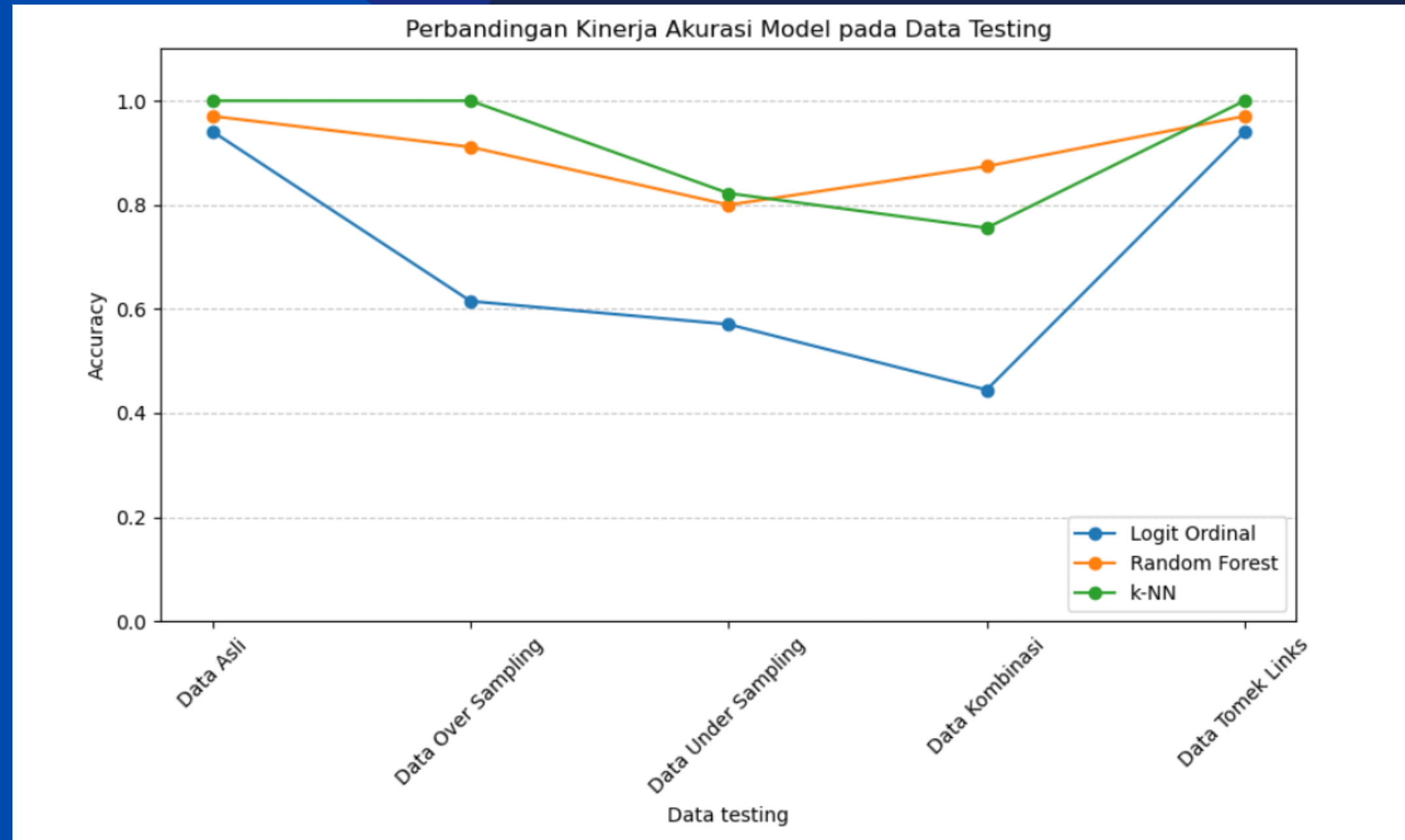
Dari pengujian pemodelan sebelumnya pasti akan didapatkan data training untuk melakukan pemodelan yang lebih rinci. untuk memperolehnya kali ini menggunakan metode logit ordinal, random forest, dan k-NN. Dengan beberapa model setelah melakukan random sampling pada dataset. Pada visualisasi tersebut dapat dilihat secara sekilas nilai yang tinggi adalah metode k-NN.

# AKURASI DATA TRAINING



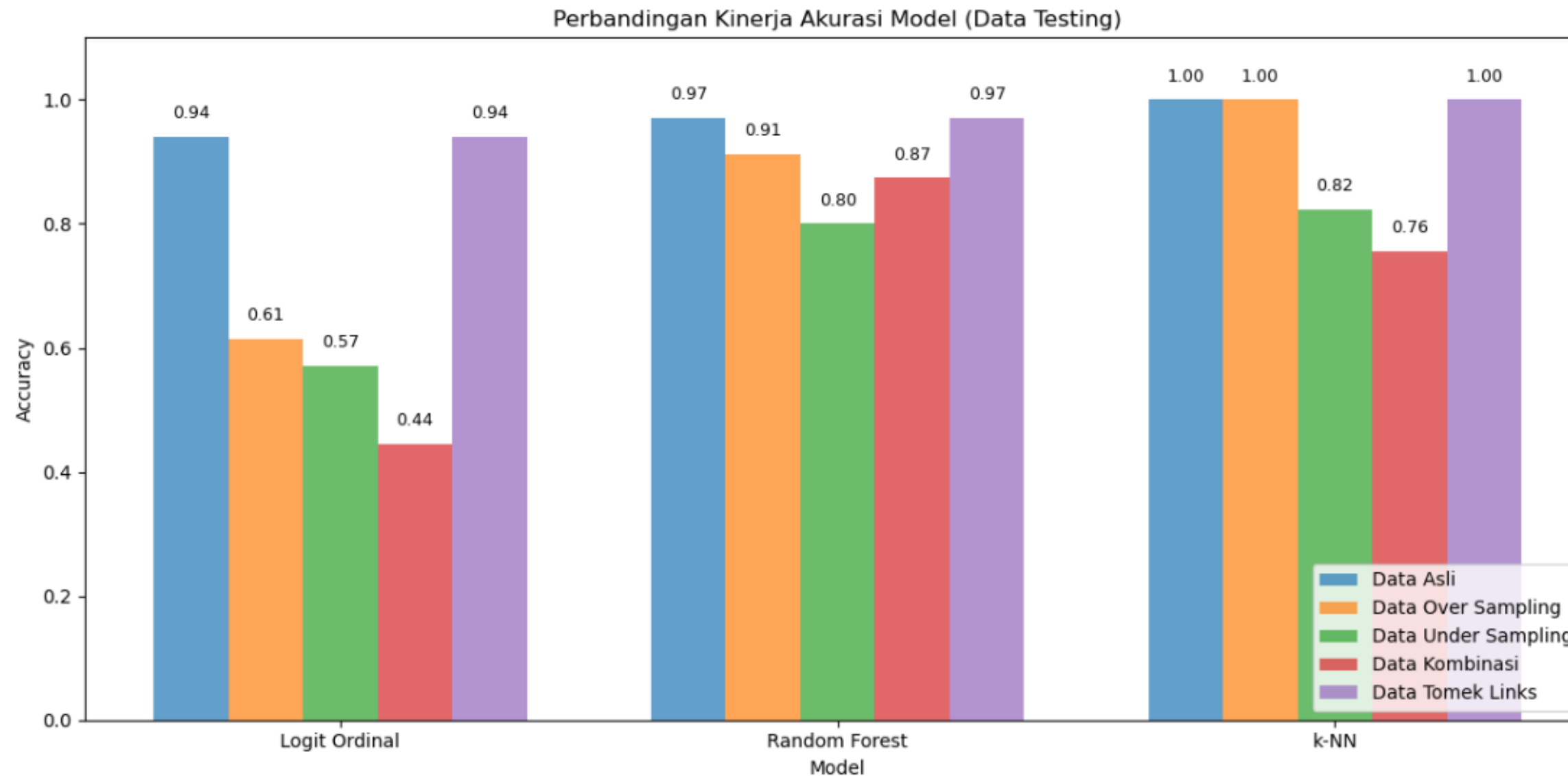
Pada visualisasi data training sebelumnya hanya dapat dilihat secara sekilas melalui grafiknya, jadi untuk memudahkan dibuatlah visualisasi tersebut dan terdapat nilai akurasi setiap model pada metodenya. Diketahui bahwa benar saja yang memiliki nilai akurasi tertinggi adalah metode k-NN yaitu semua model memiliki nilai yang sama yakni 100%. jadi metode yang terbaik pada data training adalah k-NN

# AKURASI DATA TESTING



Selain data training dalam pengujian ini juga untuk melakukan pemodelan yang lebih rinci. untuk memperoleh data testing kali ini menggunakan metode logit ordinal, random forest, dan k-NN. Dengan beberapa model setelah melakukan random sampling pada dataset. Pada visualisasi tersebut dapat dilihat secara sekilas nilai yang tinggi adalah juga metode k-NN karena grafiknya yang turun namun tetap berada diatas dibanding kedua metode yang lain

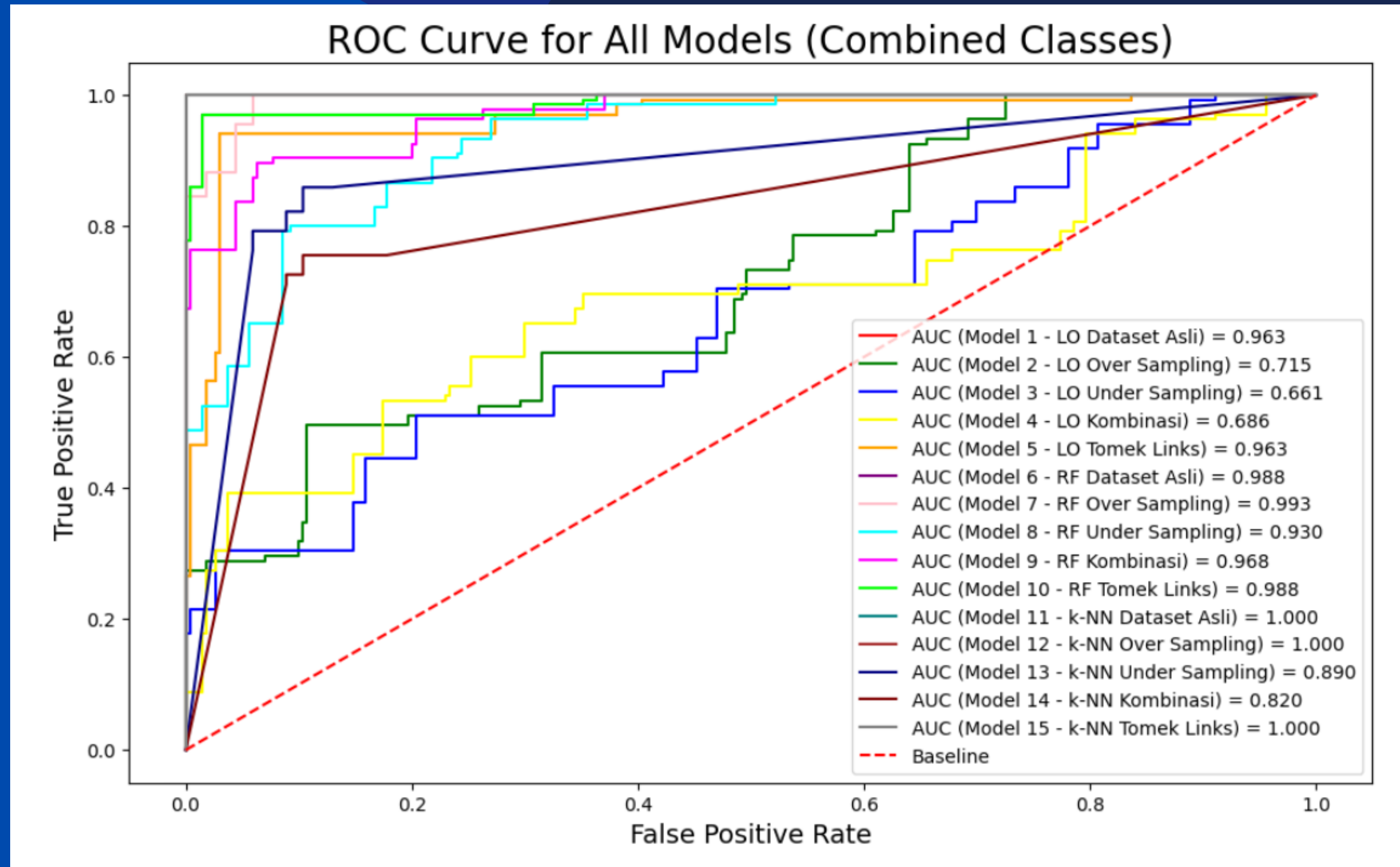
# AKURASI DATA TESTING



Pada visualisasi data training sebelumnya hanya dapat dilihat secara sekilas melalui grafiknya, jadi untuk memudahkan dibuatlah visualisasi tersebut dan terdapat nilai akurasi setiap model pada metodenya. Diketahui bahwa benar saja yang memiliki nilai akurasi tertinggi adalah metode k-NN yaitu beberapa model memiliki nilai yang sama yakni 100% dibanding metode yang lain yang hanya mendekati nilai 100% atau 1,00. jadi metode yang terbaik pada data testing adalah k-NN



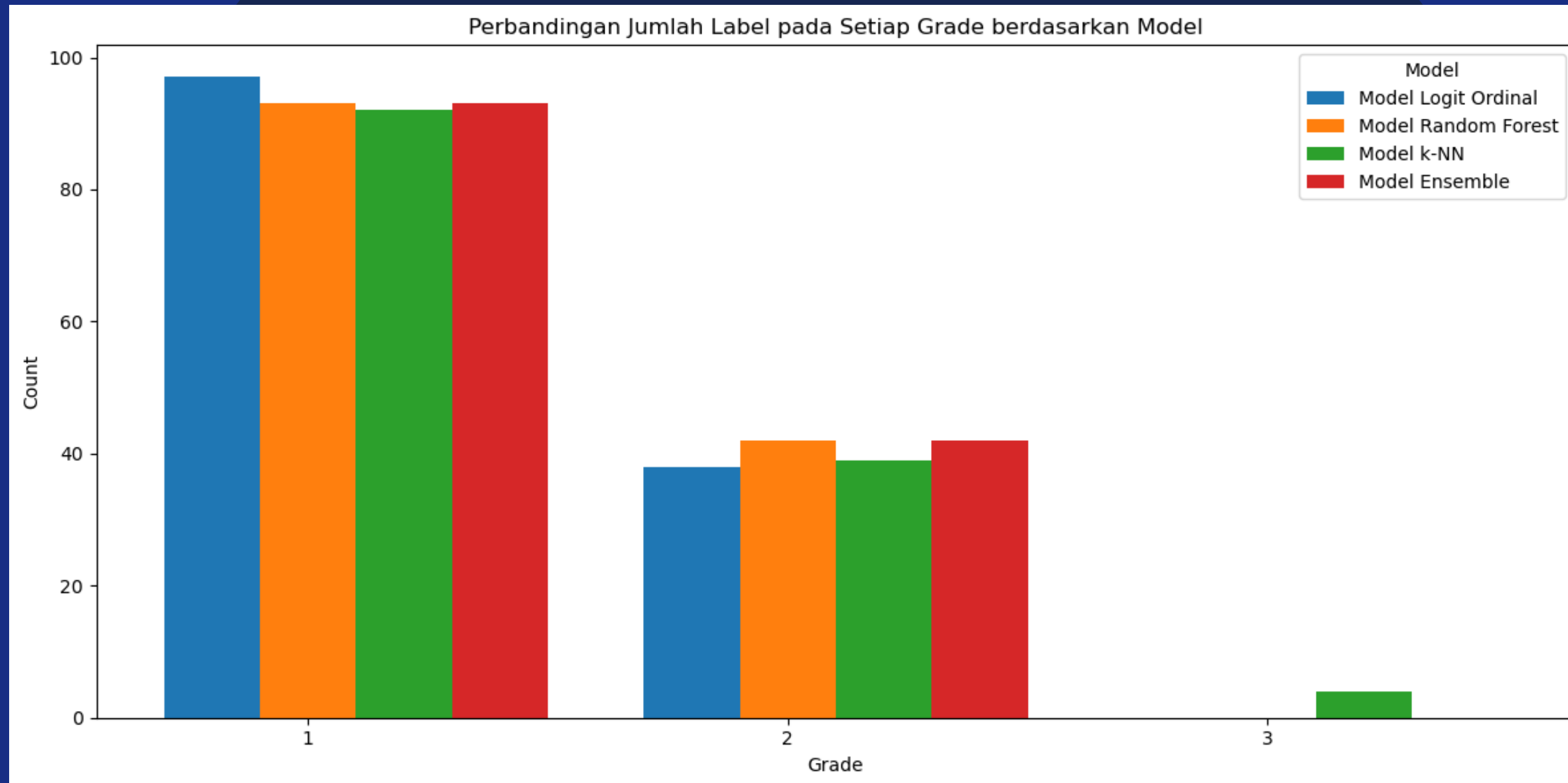
# MODEL EVALUATION



ROC yang ditampilkan adalah sebuah kurva yang didapatkan dari semua metode mulai dari logit ordinal, random forest dan k-NN. Pada nilai AUC didapatkan hasil bahwa nilai akurasi yang tertinggi adalah pada pemodelan k-NN, artinya pemodelan tersebut yang terbaik karena nilai yang dihasilkan adalah banyak yang 1,00 atau mendekati 1.



# HYBRID ENSEMBLE



Berdasarkan bar chart di atas, persebaran label untung masing-masing grade terlihat tidak merata. Grade 1 cenderung mendominasi pada semua metode, sedangkan grade 3 kebalikannya. Hal tersebut terjadi karena metode Ensemble dilakukan dengan major voting pada setiap model yang memiliki akurasi terbaik.

## KESIMPULAN

Berdasarkan pemodelan yang telah dilakukan, didapatkan bahwa metode Logit Ordinal memiliki akurasi yang stabil, sedangkan metode Random Forest dan k-NN cenderung mengalami overfitting. Setelah dievaluasi, ketiga metode tersebut dipilih untuk dilakukan ensemble pada model yang memiliki akurasi terbaik. Ensemble yang dilakukan memberikan akurasi yang tergolong tinggi, yaitu 97,037%. Walaupun demikian, metode yang memiliki akurasi terbaik adalah k-NN dengan akurasi 100% pada model11, model12, dan model 15. Di mana artinya, model-model tersebut dapat memprediksi kualitas susu secara tepat.

# THANK YOU

