Genetic Disease Risk Prediction

This presentation explores the development of a machine learning model to predict the likelihood of hereditary diseases based on genetic mutation data. We will examine the process from data collection to model evaluation and discuss ethical implications.

Understanding Hereditary Diseases

Hereditary diseases are passed down through families due to genetic mutations. These mutations can cause a wide range of health conditions, from mild to life-threatening.

Cystic Fibrosis

A condition affecting the lungs, digestive system, and other organs.

Sickle Cell Anemia

A blood disorder that affects red blood cells.

Lung Cancer

Lung cancer is a type of cancer that starts in the lungs, often caused by smoking, environmental factors, or genetic predisposition. It is a leading cause of cancer-related deaths worldwide.

Genetic Mutation Data: A Wealth of Information

Genetic mutation data provides insights into individual risk factors for hereditary diseases.

1 **Variants**

Variations in the DNA sequence, including single nucleotide polymorphisms (SNPs).

2 **Deletions**

Missing segments of DNA.

3 **Insertions**

Extra segments of DNA.

Machine Learning Models for Prediction

Machine learning models can analyze genetic mutation data to predict the likelihood of developing a hereditary disease.

Decision Trees

Simple models that create a series of decisions based on features.

Random Forest

Random Forest combines multiple decision trees for improved accuracy and robustness

Multinominal Model

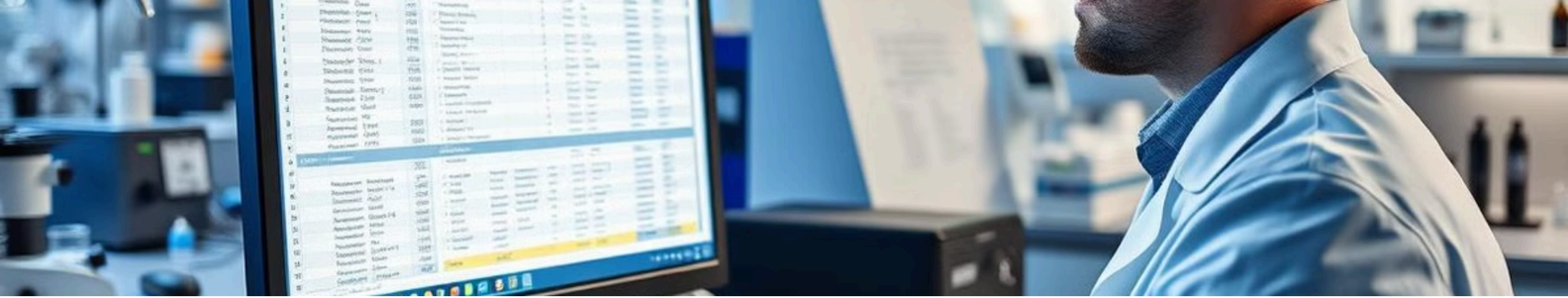
A multinomial model extends logistic regression to handle multi-class classification.

Naive Bayes

Naive Bayes is a probabilistic classifier based on Bayes' theorem and feature independence.

Logistic Regression

Logistic regression is a statistical model used for binary classification, where the outcome is a probability that maps to two possible classes.



Feature Engineering and Data Preprocessing

Transforming raw genetic data into meaningful features for the model.

1

Data Cleaning

Removing inconsistencies and errors.

2

Feature Selection

Choosing the most relevant features for prediction.

3

Data Normalization

Scaling data to a consistent range.

4

Accuracy Prediction

predicting accuracy of the model.

Model Selection and Tuning

Choosing the best model for the task and optimizing its parameters.

1

Cross-Validation

Splitting the data into training and testing sets.

2

Hyperparameter Optimization

Adjusting model settings to improve performance.

3

Model Evaluation

Assessing the model's ability to make accurate predictions.

Evaluating Model Performance

Metrics for evaluating model accuracy and reliability.

91%

Accuracy(Random Forest)

Percentage of correct predictions.

0.85

Precision

Ability to identify true positives.

0.92

Recall

Ability to identify all positive cases.

Ethical Considerations and Future Directions

It is crucial to address ethical concerns related to genetic risk prediction and ensure responsible use of the technology.

1

Privacy

Protecting sensitive genetic information.

2

Discrimination

Preventing biased use of predictions.

3

Access and Equity

Ensuring equitable access to testing and predictive services.