# Empowering Nile's Marketing :
# Predictive Modeling for Customer Reviews

Created by:
5663833, 5587796, 5630106, 5663354, 5635111, 5595277

# Business Understanding

**1**

### Develop Predictive Model

Create a machine-learning model to identify customers likely to leave positive reviews

**2**

### Improve Online Reputation

Enhance Nile's reputation through increased positive feedback

**3**

### Drive Sales Growth

Leverage improved reputation to increase sales and market share

**4**

### Cost-Effective Solution

Implement a model that provides valuable insights without excessive resource allocation

# Data Understanding: Available Datasets

**Reviews**

- Order_ID
- Review_answer_timestamp
- Review_score

**Order**

- Order_ID
- Order_status
- Order_purchase_timestamp
- Delivery_customer_date
- Order_estimated_delivery_date

**Items**

- Order_ID
- Frieght_value
- Price

# Data Preparation

**1** **Data Cleaning**

Remove missing values and correct data types

**2** **Data Filtering**

Focus on accurate and reliable data for analysis

**3** **Feature Engineering**

Create new columns to enhance predictive power

**4** **Outlier Removal**

Identify and remove extreme values to improve model reliability

# Data Preparation

| New columns | Meaning |
|---|---|
| Total Value | The total cost of each order |
| Freight_percentage | The proportion of freight cost in each order |
| On_time | The punctuality of the order |
| Delivery_time | Delivery time of each order in days |

# Exploratory Data Analysis

**Distribution of Review Counts Across Order Statuses**
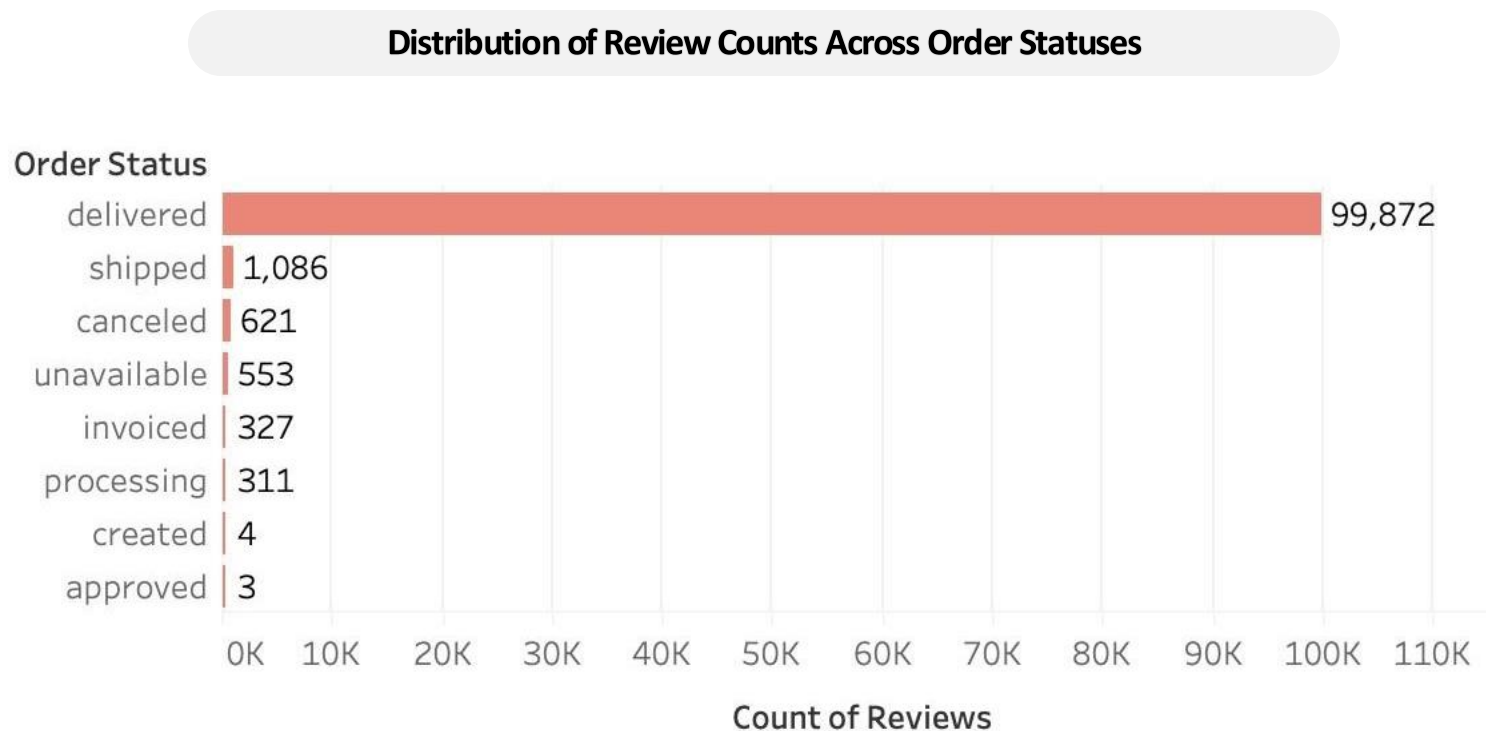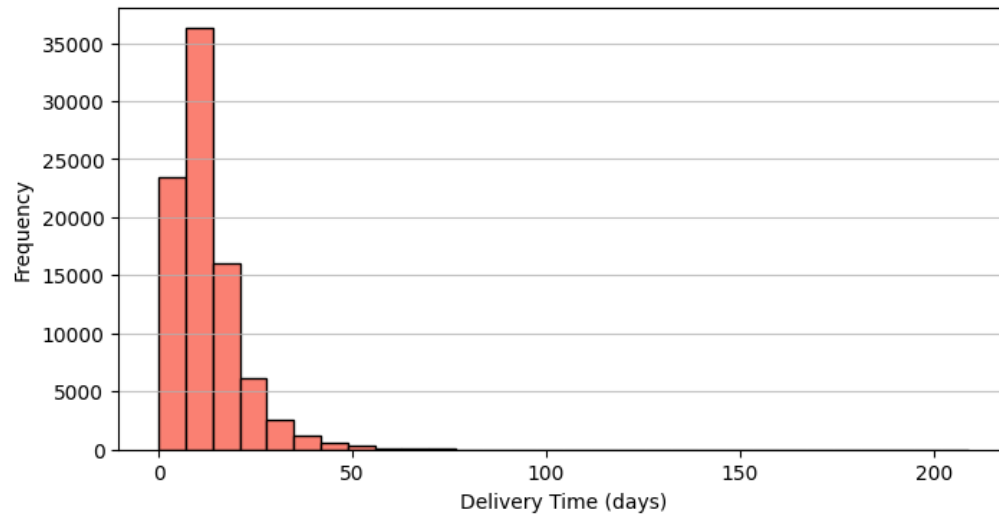


- Most orders in our dataset are 'Delivered', providing a solid foundation for analysis
- Only focus on 'Delivered' orders to target completed transactions and their reviews
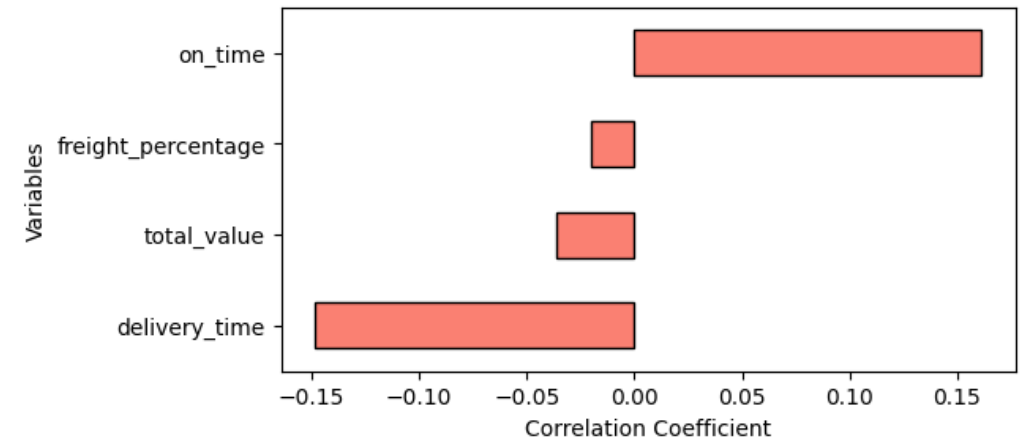
# Exploratory Data Analysis
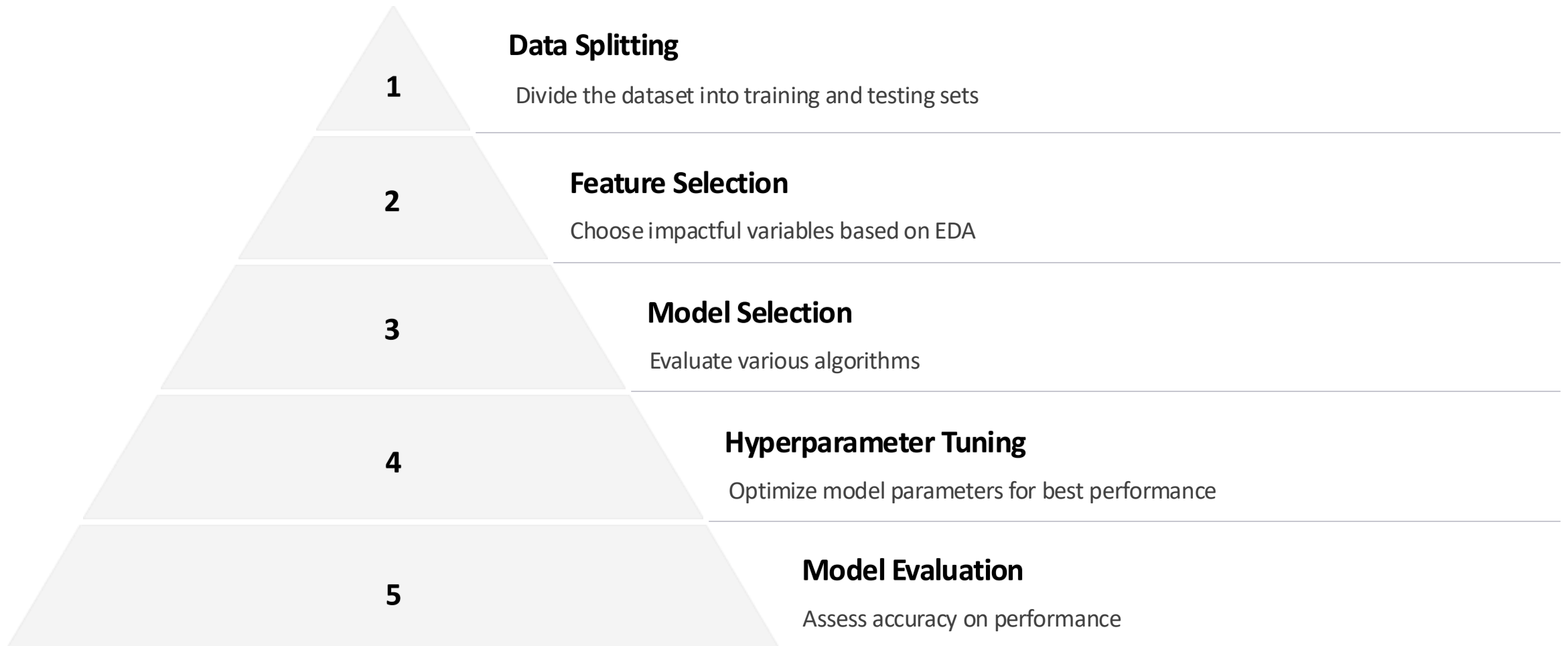
### Distribution of Delivery Time



### Relationship with Review Score



- Outliers identified using a delivery time histogram

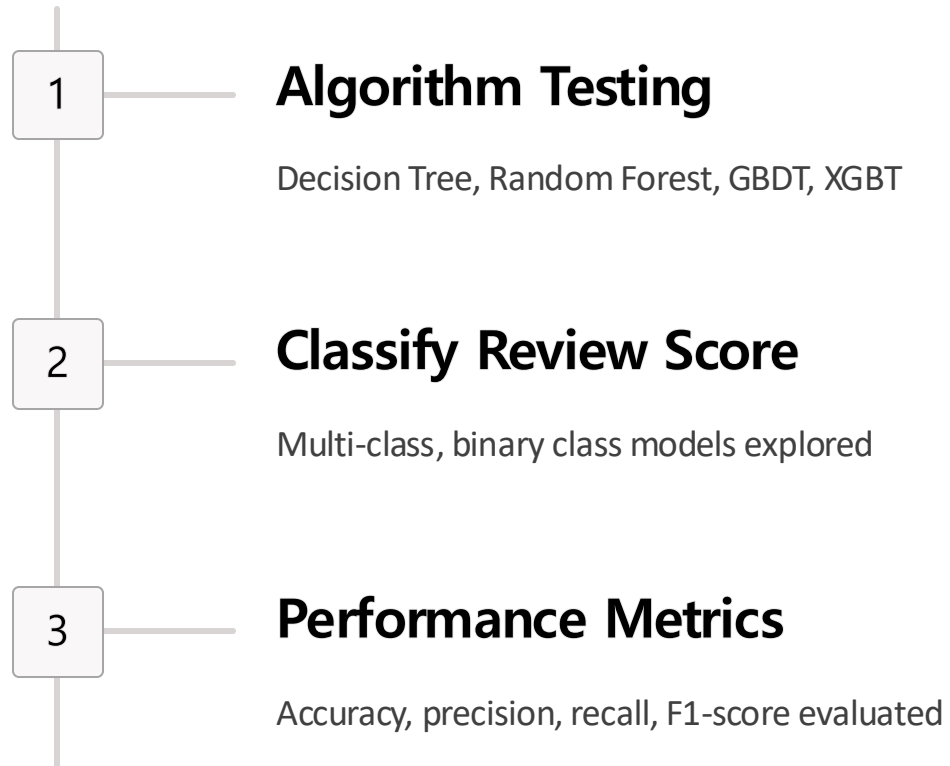- Extreme delivery times removed to maintain data quality

- **On-time Delivery**: Positive relationship with review scores

- **Delivery Time**: Negative impact on customer satisfaction

- **Total Value & Freight Percentage**: Minimal connection

# Modelling Process

**Data Splitting**

1

Divide the dataset into training and testing sets

**Feature Selection**

2

Choose impactful variables based on EDA

**Model Selection**

3

Evaluate various algorithms

**Hyperparameter Tuning**

4

Optimize model parameters for best performance

**Model Evaluation**

5

Assess accuracy on performance

# Model Selection Process

**1** **Algorithm Testing**

Decision Tree, Random Forest, GBDT, XGBT

**2** **Classify Review Score**

Multi-class, binary class models explored

**3** **Performance Metrics**

Accuracy, precision, recall, F1-score evaluated

# Model Performance Results

**Two review classifications outperformed**

Highest accuracy at 82% with GBDT and XGBT

**GBDT Model Selected**

Balanced performance across key metrics

**Underpredicting smaller groups**

Challenge in identifying low review scores

# Model Limitations

| What limited model performance? | |
|---|---|
| **Features used** | Basic features that limited strong performance |
| **Data filtering** | Narrowing down the data to make the dataset overclean |
| **Fewer relationships** | Fewer impactful variables for machine learning |
| **Imbalanced review data** | The model struggled to predict minority class |

# Deployment Plan
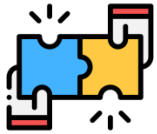
## System Integration

Integrate with existing CRM and ERP systems

for seamless data exchange

## Team Training

Conduct sessions for marketing and customer service teams

## Performance Monitoring

Implement systems to track model performance and con

duct regular maintenance
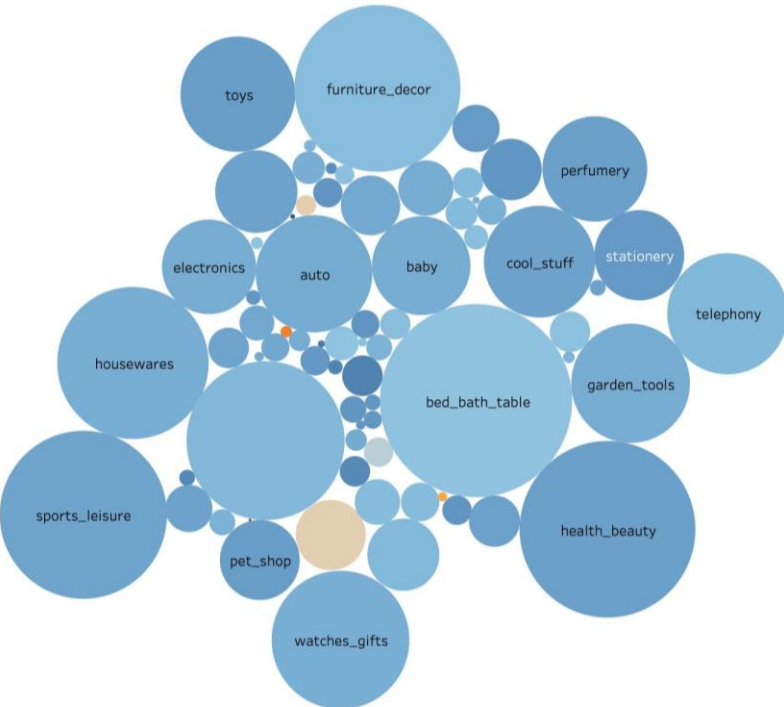
## Privacy and Ethics

Ensure compliance with data protection regulations and add

ress ethical concerns

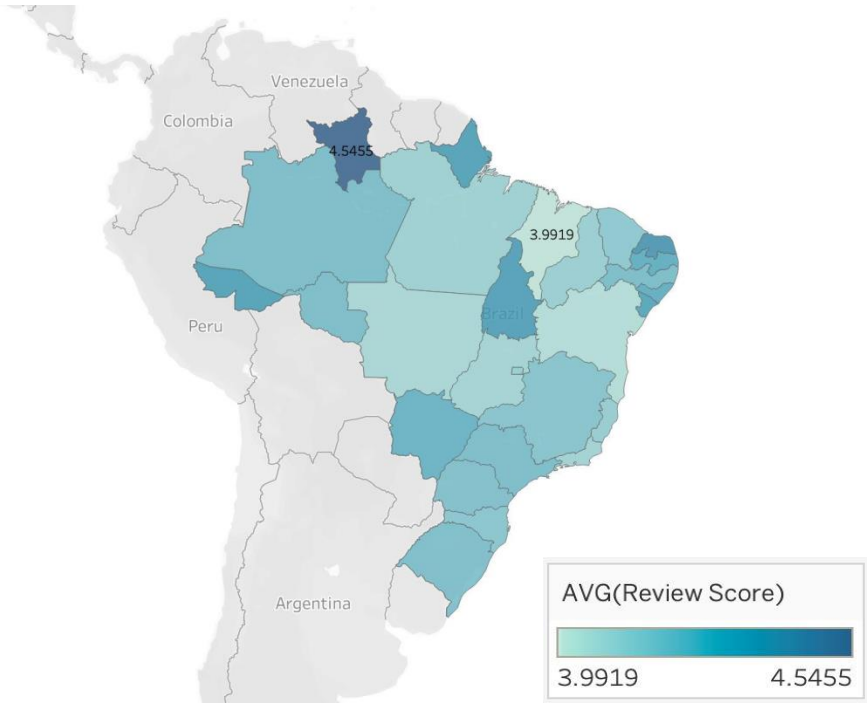# Suggestions for Next Steps
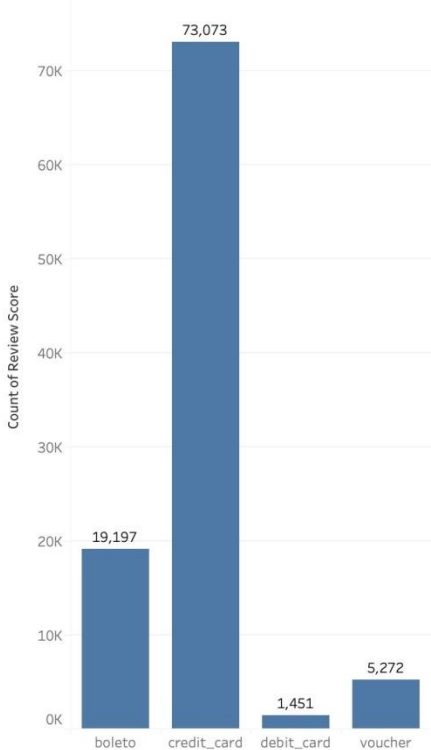
### 1   Create more features



**Average Review Score by Product Category**

**Customer Satisfaction Map Across Brazil**

**Distribution of Review Counts across Payment Methods**

# Suggestions for Next Steps

1 **Create more features**

More impactful features extracted from existing dataset s (i.e. product category, payment type)

2 **Data expansion**

Collect and consider more data related to revie ws for increasing model complexity

3 **Balance the review data**

Increase negative reviews to balance two classes for pre dicting

4 **Implement multiple models**

Combine different algorithms for better results (i.e. st acking & voting)

Thank you