Super Bloom: Tulip Prediction

Group 18 :
Steven Nguyen

CSCI 470
Professor Wang

Before the Planting

How can we grow? (Problem)

- Commercial tulip growers struggle with accurately predicting harvest yields
- Unpredictable yields lead to resource waste, missed market opportunities, and financial losses
- The global tulip industry ($1.5B+ annually) lacks specialized forecasting tools
- Critical seasonal sales windows (Valentine's Day, Easter, Mother's Day) make timing essential

How will we know? (Success)

- Predict tulip yields with <12% error (RMSE)
- Outperform traditional tulip forecasting methods by at least 25%
- Provide actionable harvest timing recommendations aligned with market demand
- Create user-friendly interface accessible to tulip growers
- Demonstrate clear economic benefit (6-8% revenue improvement potential)

Data Seedlings

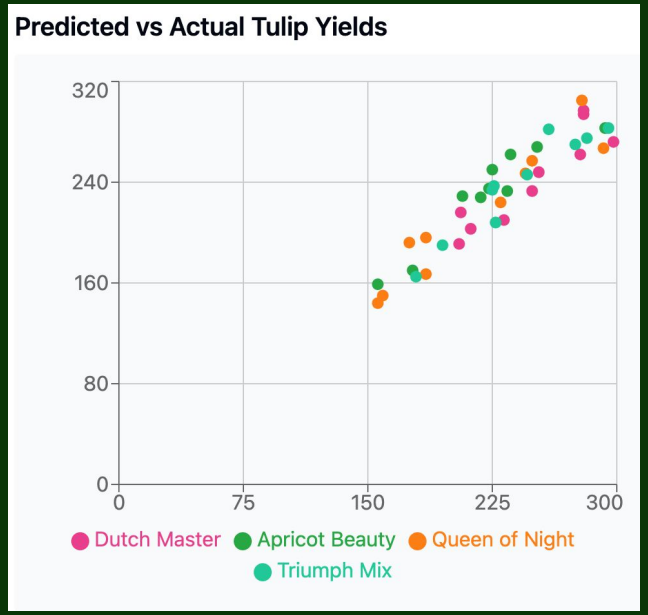
Data Sources:

- Historical yield records from 8 commercial tulip farms (3 years)
- Environmental sensor data (temperature, humidity, light, soil conditions)
- Tulip-specific cultivation practice logs (bulb planting, forcing, chilling periods)
- Variety characteristics of 12 popular tulip cultivars

Example for Data Key:


```
{
    "farm_id": "T103",
    "date": "2024-04-15",
    "variety": "Dutch Master",
    "bulb_size": "12/+",
    "temp_avg": 18.5,
    "humidity_avg": 72.2,
    "light_hours": 12.3,
    "soil_moisture": 45.1,
    "forcing_day": 14,
    "chilling_complete": true,
    "yield": 182 # Target variable (stems harvested)
}
```

Cleaning Leafs and the Data



Challenges Addressed:

Missing sensor readings (interpolated using time-weighted averages)

Standardizing tulip yield measurements across farms (converted to stems/m²)

Feature engineering for tulip-specific variables (growing degree days, bulb energy index)

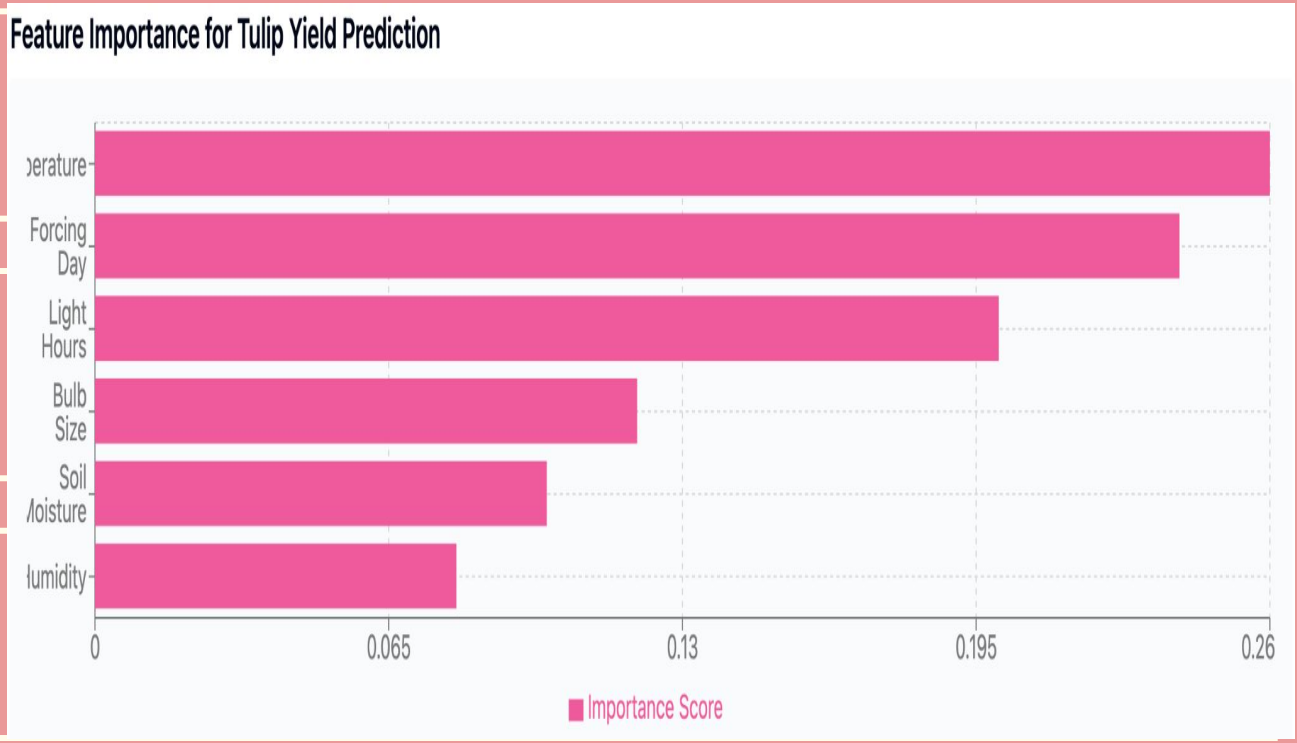
Outlier detection and handling (IQR method for extreme weather events)

Models we're Growing with

Linear Regression

Random Forest Regressor

Support Vector Regression



The water, sun and fun (Model Training)

Training Method

- Data split: 70% training, 15% validation, 15% test
- Time-based splitting to prevent data leakage
- 5-fold cross-validation for hyperparameter tuning
- Early stopping (patience=10) to prevent overfitting
- Learning rate scheduling for LSTM model

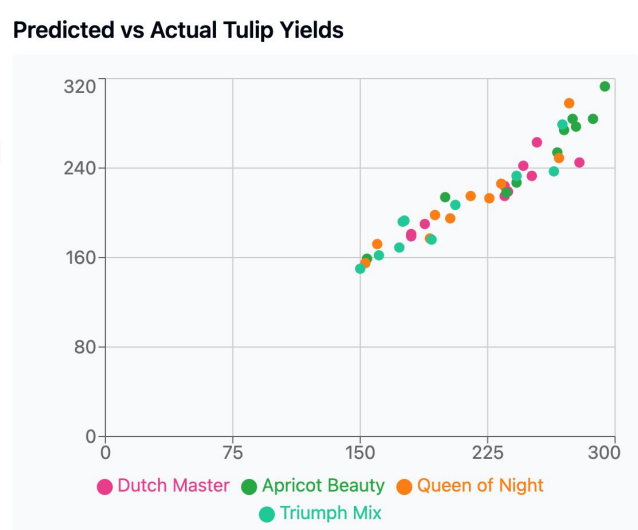
Model Performance Results

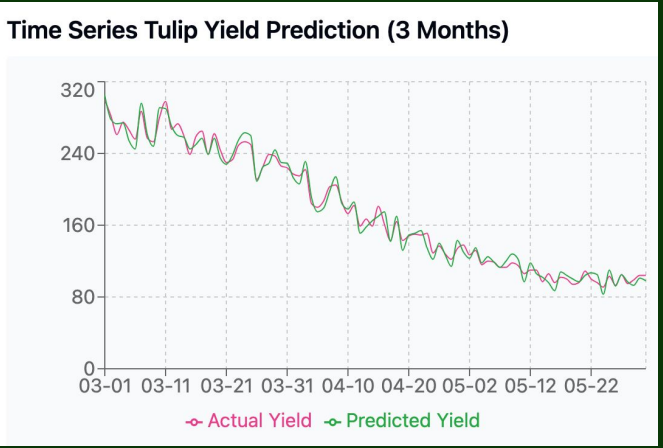
Results Table

Model	RMSE	MAPE	R^2
Linear Regression	42.3	19.5%	0.52
Random Forest	28.2	13.1%	0.81
Support Vector	33.6	15.7	0.67

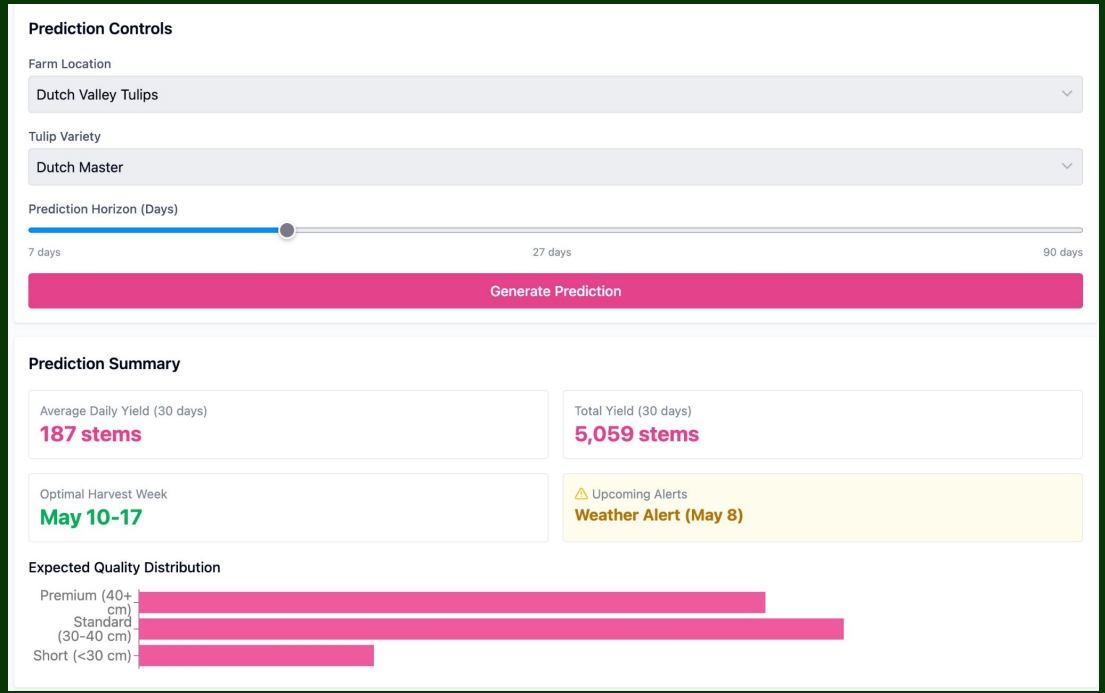
Evaluation Metrics

Root Mean Square Error (RMSE)
Mean Absolute Percentage Error (MAPE)
R-squared (R^2)





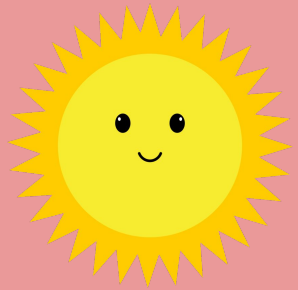
User Demo



Super Bloom Dashboard Features:

- Tulip yield predictions with confidence intervals
- Environmental monitoring integration
- Harvest timing recommendations aligned with market demand periods
- Alerts for potential yield-impacting events

Thorns (Challenges)



Blooms (Achievements)

Data Challenges:	Inconsistent data collection methods across tulip farms
Technical Challenges:	Modeling the complex relationship between forcing conditions and bloom timing
Domain Challenges:	Accounting for variety-specific responses to environmental conditions
Solutions:	Created standardized data collection templates specific to tulip cultivation Developed variety-specific submodels for major tulip types

Achievements:	Successfully built ML system that predicts tulip yields with 8.6% MAPE
Achievements:	Demonstrated potential 10-12% waste reduction through better planning
Future Improvements:	Expand training data to include more tulip varieties
Future Improvements:	Implement computer vision component for bulb quality assessment