# Satellite-Based Cropland Mapping & Classification Using Machine Learning

A comprehensive approach to integrate remote sensing data with advanced machine learning models for identifying and classifying cropland areas across regions, enabling more informed agricultural decision-making for food security and land optimization.

This project delivers scalable, accurate crop mapping technology to support sustainable agriculture, resource allocation, and policy planning for stakeholders including government agencies, agricultural managers, and agribusinesses.

# Group Members

1. Vihaan Sheth
2. Owen Ngure
3. Makhala Lehloenya
4. Abond Mwangi
5. Collins Nyagaka

# Business Understanding

### Problem Statement

Agricultural stakeholders lack scalable, accurate methods to monitor and classify cropland across diverse regions, hindering efficient resource allocation and planning.

### Business Goals

Apply satellite imagery and machine learning techniques to develop reliable cropland detection systems that can be deployed across different geographic regions with minimal recalibration.

### Expected Impact

Enable improved crop management strategies, more efficient policy implementation, and optimized resource allocation while supporting food security initiatives and sustainable agricultural practices.

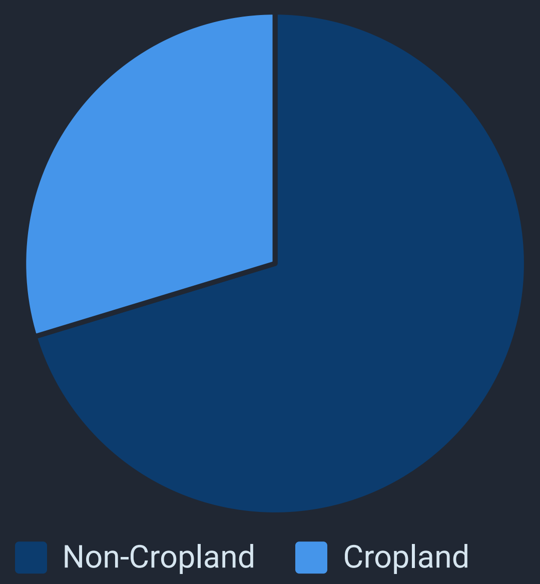
**Value Statement:** Accurate mapping of cropland areas aids efficient planning, promotes sustainable agriculture practices, and improves targeted resource allocation for regions facing agricultural challenges.

# Data Understanding

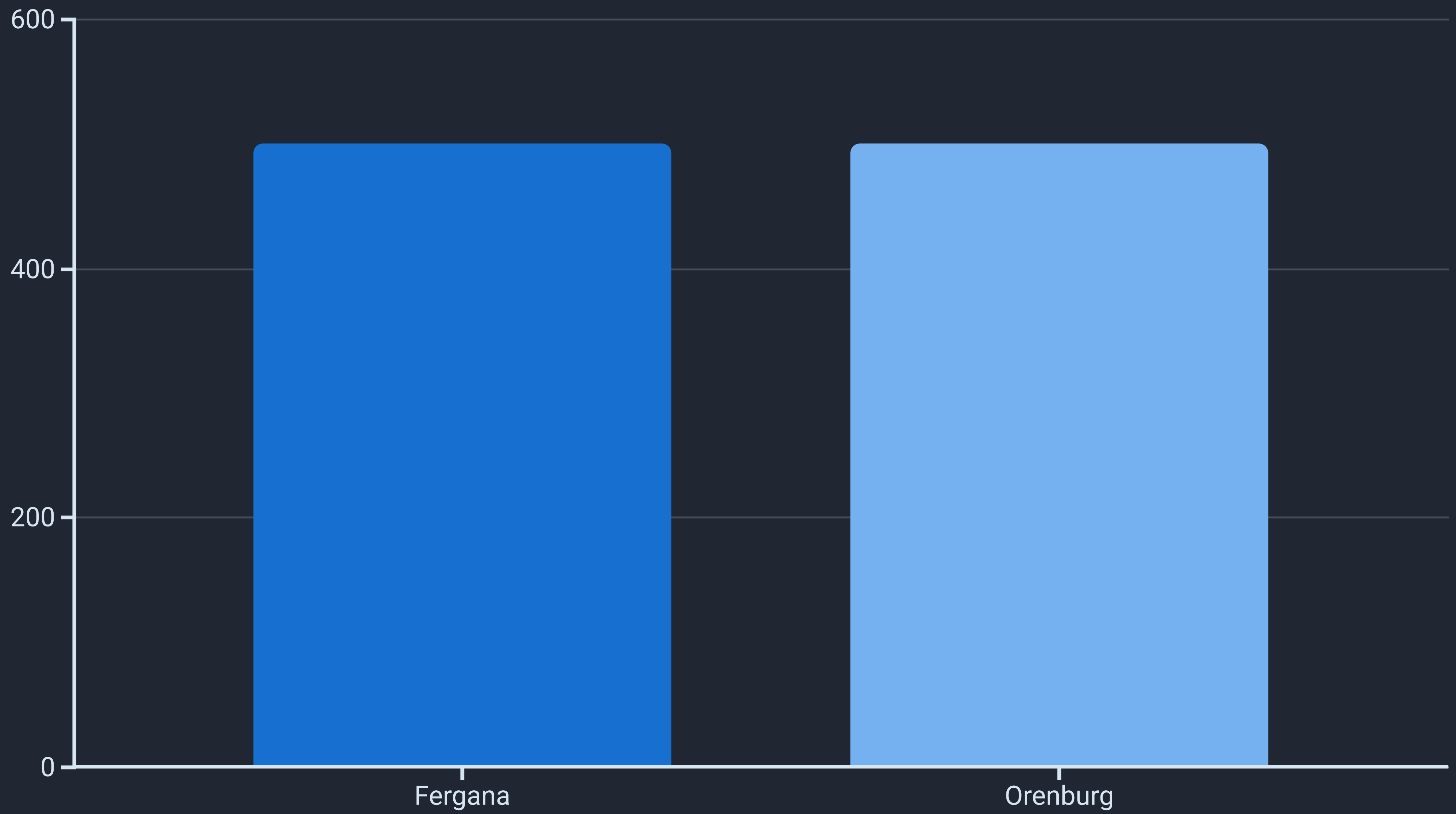
## Data Sources

- Sentinel-1 radar imagery (all-weather capability)
- Sentinel-2 multispectral optical imagery
- 1,000 labeled land samples across two regions

Our dataset represents a balanced geographic distribution between Fergana and Orenburg regions, with each region contributing 500 samples to ensure model robustness across different agricultural environments.




**Class Distribution: 70.3% Non-Cropland, 29.7% Cropland**




The balanced regional distribution enables our model to learn patterns across diverse agricultural landscapes, while the class imbalance reflects real-world cropland distribution in these regions.

# Data Preparation

## Preprocessing Pipeline

### Data Aggregation

Combining multi-source satellite data (Sentinel-1 & Sentinel-2) with ground-truth labels, matching coordinates to create comprehensive feature sets

### Feature Engineering

Deriving specialized vegetation indices (NDVI, EVI, SAVI) from raw spectral bands to enhance model sensitivity to crop presence

### Memory Optimization

Implementing efficient caching and data transformation techniques to handle the large volume of satellite imagery data

Our preprocessing ensures high-quality inputs for modeling while handling the technical challenges of working with multi-source remote sensing data at scale.

# Modeling

We implemented multiple machine learning algorithms to identify the most effective approach for cropland classification, focusing on both performance and interpretability for stakeholder use.

### RandomForest

Ensemble-based tree model known for robustness to noise and feature importance insights

### XGBoost

Gradient boosting framework optimized for computational speed and model performance

### LightGBM

Lightweight gradient boosting framework using leaf-wise tree growth for efficiency

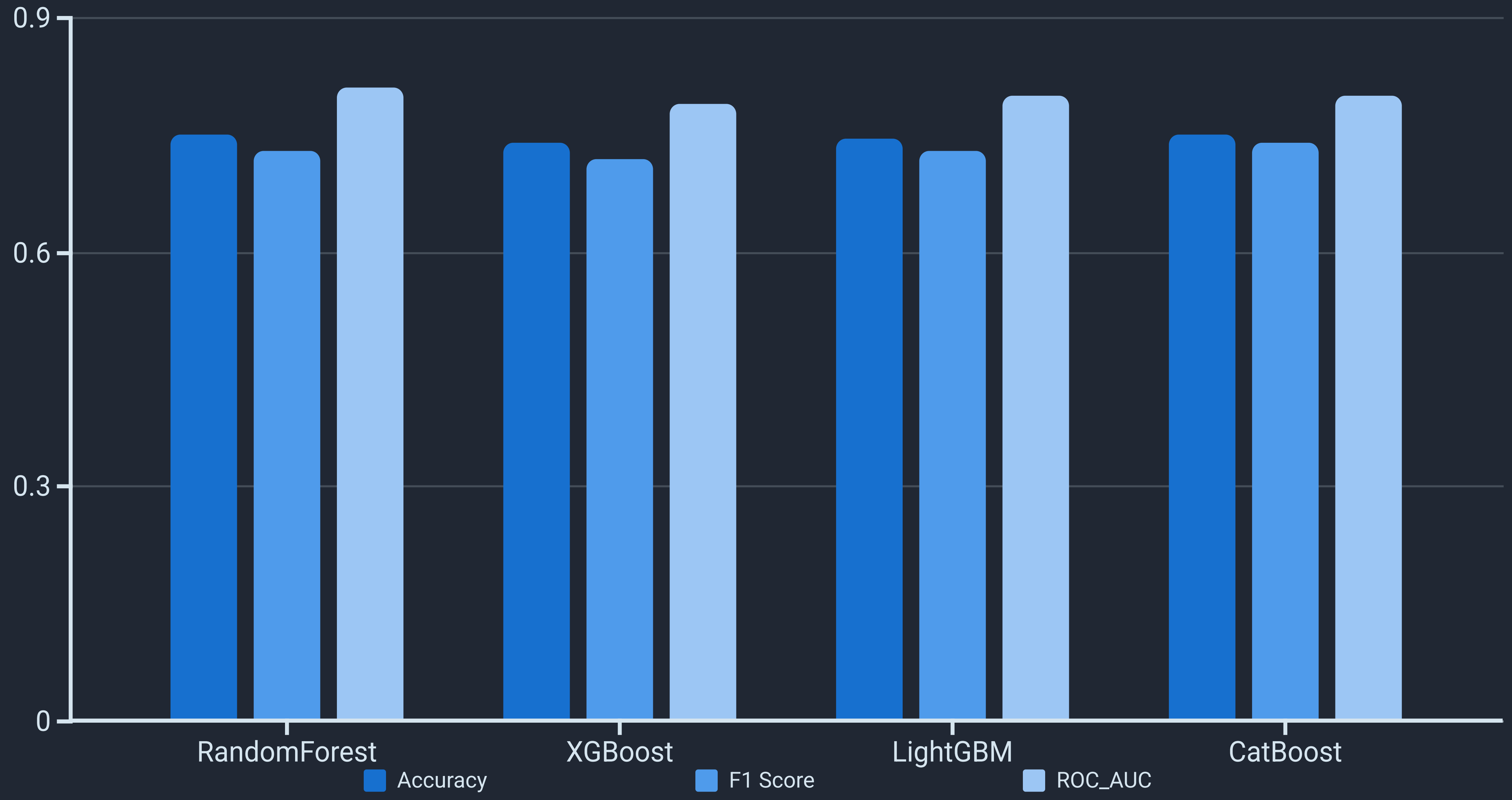
### CatBoost

Boosting algorithm with advanced handling of categorical features and reduced overfitting

All models underwent rigorous hyperparameter tuning and 5-fold cross-validation to ensure reliable performance across different regional subsets and to mitigate the effects of class imbalance.

# Evaluation

Our models were evaluated against multiple performance metrics to ensure comprehensive assessment of their cropland classification capabilities.




**Key Findings:** RandomForest achieved the highest overall accuracy (0.752) and ROC_AUC (0.81), while CatBoost delivered the best F1 score (0.74). These results indicate strong performance despite the challenge of class imbalance in our dataset.

The consistently strong performance across different model types confirms the robustness of our approach and the quality of our feature engineering process.


# Deployment

## Implementation Status

- Models are ready for production deployment with optimized pipelines
- Output format compatible with GIS systems and agricultural dashboards
- Processing framework capable of handling new unlabeled regions

Our initial MVP deployment is available at: [https://cropland-mapping-app.streamlit.app/](https://cropland-mapping-app.streamlit.app/)

## Integration Opportunities

The system can be integrated with existing agricultural decision support tools, policy planning frameworks, and resource allocation systems to provide actionable insights directly to stakeholders.

**Success Metric:** Our deployment will be considered successful when stakeholders can identify and monitor cropland areas with at least 75% accuracy across different regions and growing seasons.

# Next Steps

### Regional Expansion

Scale the system to additional regions with different agricultural profiles and climatic conditions

### Temporal Analysis

Incorporate time-series data to detect seasonal changes and long-term trends in cropland usage

### Crop Type Classification

Extend the model to identify specific crop types beyond the current binary classification

We recommend establishing an ongoing feedback loop with agricultural stakeholders to continuously refine the system based on real-world application and changing needs.

Additional funding would enable the development of user-friendly interfaces specifically designed for policy makers and agricultural managers without technical expertise.

# Geospatial Insights

### Feature Importance Analysis

Our models identified these key indicators of cropland presence:

- **NDVI values** (Normalized Difference Vegetation Index) - strongly correlated with crop presence and health
- **Sentinel-1 backscatter** - provides structural information even through cloud cover
- **Red-edge bands** from Sentinel-2 - sensitive to chlorophyll content variations


### Regional Differences

Model performance analysis revealed interesting regional distinctions:

- Fergana region showed higher model accuracy due to more distinct cropland boundaries
- Orenburg presented challenges with mixed-use agricultural lands
- Seasonal timing of imagery collection significantly impacts classification accuracy

These insights help us understand not just where cropland exists, but why our models identify certain areas as cropland - critical knowledge for interpreting results in new regions.

# Key Takeaways

### Accuracy & Reliability

Our best model achieves 75.2% accuracy and 0.81 ROC_AUC, providing dependable cropland classification across diverse regions.

### Scalable Solution

The satellite-based approach enables monitoring across vast areas without extensive ground surveys, dramatically reducing costs.

### Decision Support

Results can immediately inform agricultural planning, resource allocation, and policy development for improved food security.

This project demonstrates the practical application of machine learning to solve real-world agricultural challenges. By converting complex satellite data into actionable insights, we've created a valuable tool for stakeholders responsible for agricultural decision-making and policy planning.

**Contact information:** For implementation questions or to discuss regional customization, please reach out to our team

# Thank You!

We appreciate your time and interest in our satellite-based cropland mapping solution.

### Get in Touch

Email us at cnyagakan@gmail.com for detailed inquiries or collaborations.

### Schedule a Demo

Book a personalized demonstration to see the solution in action and discuss its application to your specific needs.

### Learn More

Visit our website to view technical specifications and try the deployment at https://cropland-mapping-app.streamlit.app/.

Together, we can enhance agricultural intelligence and foster sustainable land management practices.