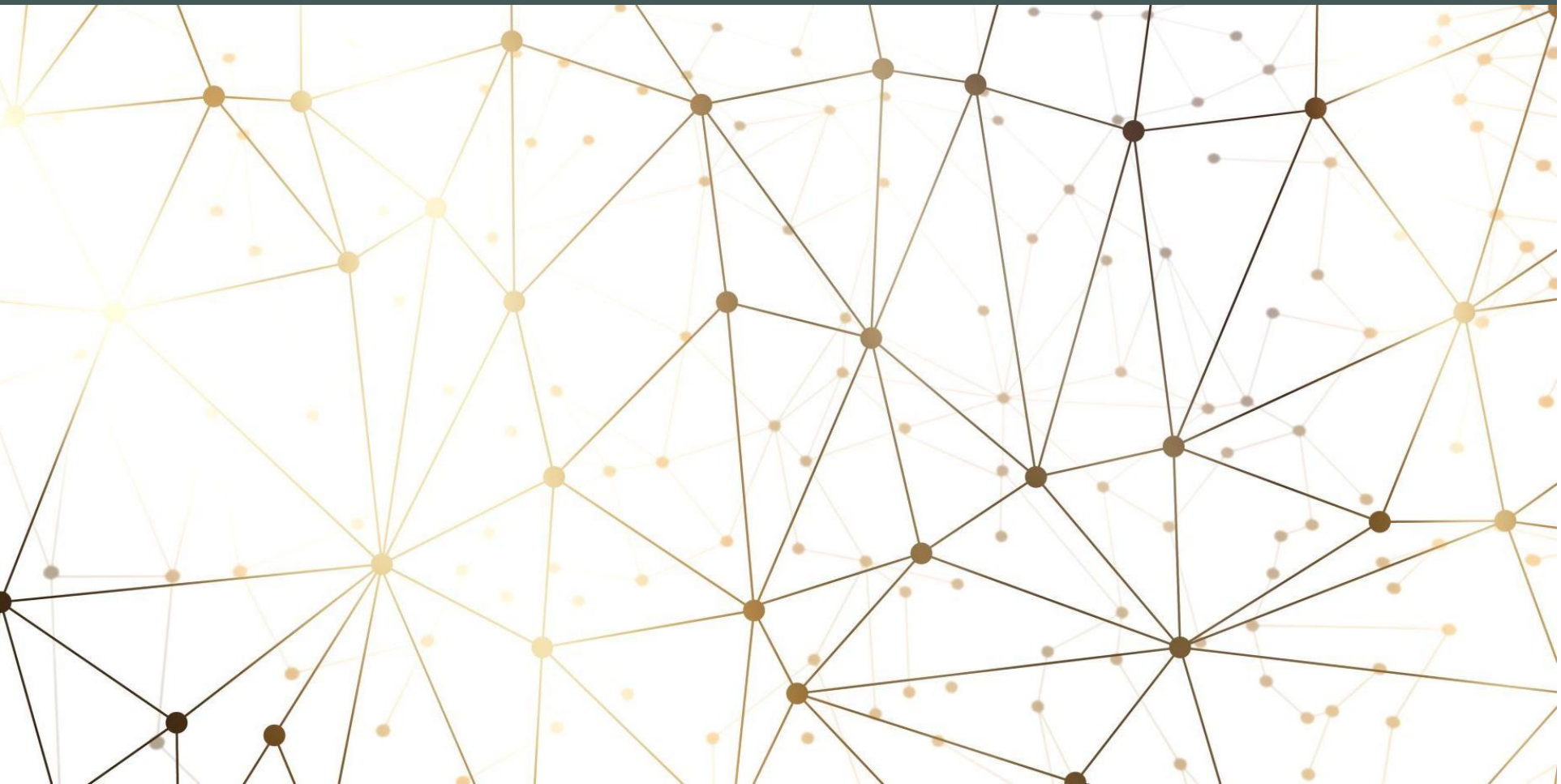
AI/ML-Based Predictive Analysis for Kubernetes Cluster Failures

Proactive Monitoring and Issue Prevention



Presented By
1. Arati Balaji
2. Preetika Pradeep
3. Satyavan

Problem Statement

- Kubernetes clusters face failures like:
  - Pod crashes
  - Resource bottlenecks
  - Network issues

Goal : Predict these failures before they occur using AI/ML models.

Solution Approach

- Utilize historical and real-time cluster metrics.
- Train machine learning models for predictive analysis.
- Deploy a proactive monitoring system and provding a UI which is easy to use and access.

Data Collection & Preprocessing

- Data sources: Kubernetes metrics, logs, and resource utilization.
- Preprocessing steps:
  - Handling missing values
  - Feature selection & engineering
  - Data normalization

Models Trained

- Decision Tree Classifier
- Random Forest Classifier
- XGBoost Classifier
- Stacking Classifier (combining multiple models)

Model Training & Evaluation

- Models trained on historical Kubernetes cluster data.
- Evaluation using accuracy scores and confusion matrices.

Model Performance Comparison

- Decision Tree Classifier: **0.9995**
- Random Forest: **0.6165**
- XGBoost: **Confusion Matrix Output**
- Stacking Algorithm: **1.0**

Deployment Strategy

- Model integrated with Gradio UI for user-friendly interaction.
- Real-time inference with an intuitive dashboard.
- Visual representation of predictions for easy monitoring.

Future Enhancements

- Improve dataset diversity for better generalization.
- Implement reinforcement learning for adaptive tuning.
- Enhance model interpretability for better decision-making.

Conclusion

- AI/ML models significantly improve Kubernetes failure predictions.
- Stacking and Decision Tree models show the best performance.
- Deployment of predictive monitoring enhances system reliability.