Update to Advisory Council

Presenter: Lily Zhuhadar

| Director,
Center for Applied Data Analytics
| Co-Director,
Data Science Program

- **Email:** lily.popova.zhuhadar@wku.edu
- **Cellphone:** (615) 604-4995
- **Office Address**
  WKU Gordon Ford College of Business
  410 Regents Avenue
  CHAN Building, Room 3049
  Bowling Green, KY 42101

WKU GORDON FORD COLLEGE OF BUSINESS

Center For Applied Data Analytics

DIRECTOR: LILY POPOVA ZHUHADAR, PH.D.

CONNECTING RESEARCHING INNOVATING

September 19, 2025

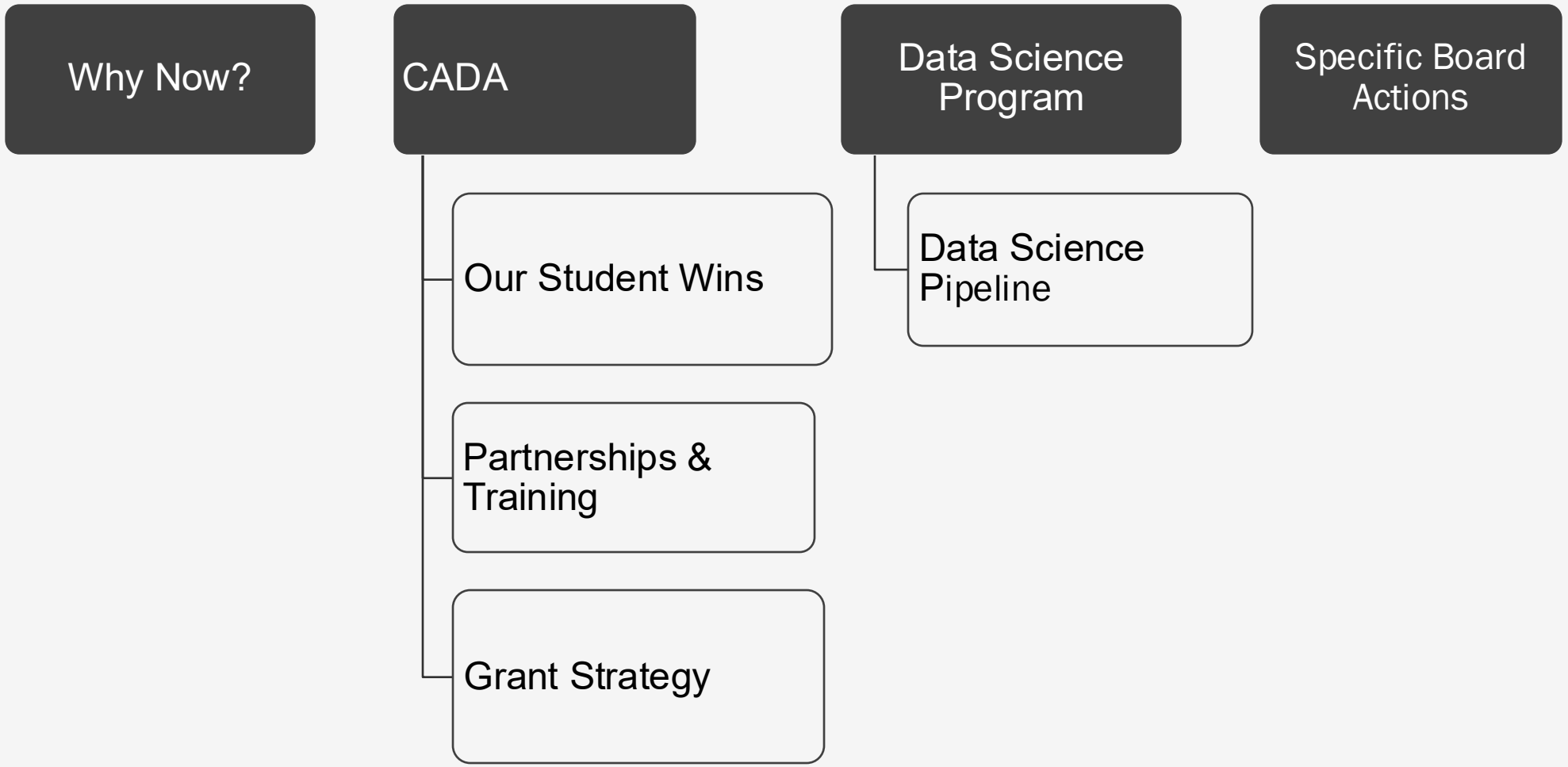
Update to Advisory Council

Presenter: Lily Zhuhadar

| Director,
Center for Applied Data Analytics
| Co-Director,
Data Science Program

Download this Presentation

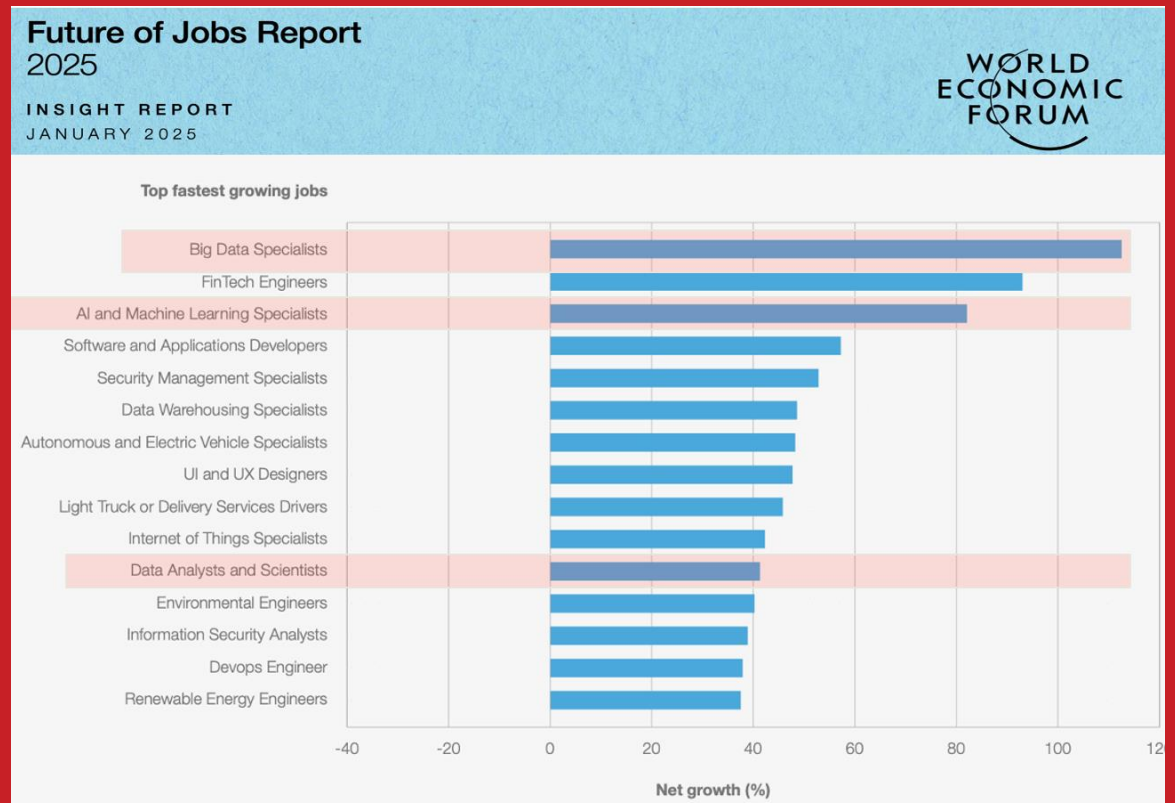
Game Plan for the Next 15 Minutes

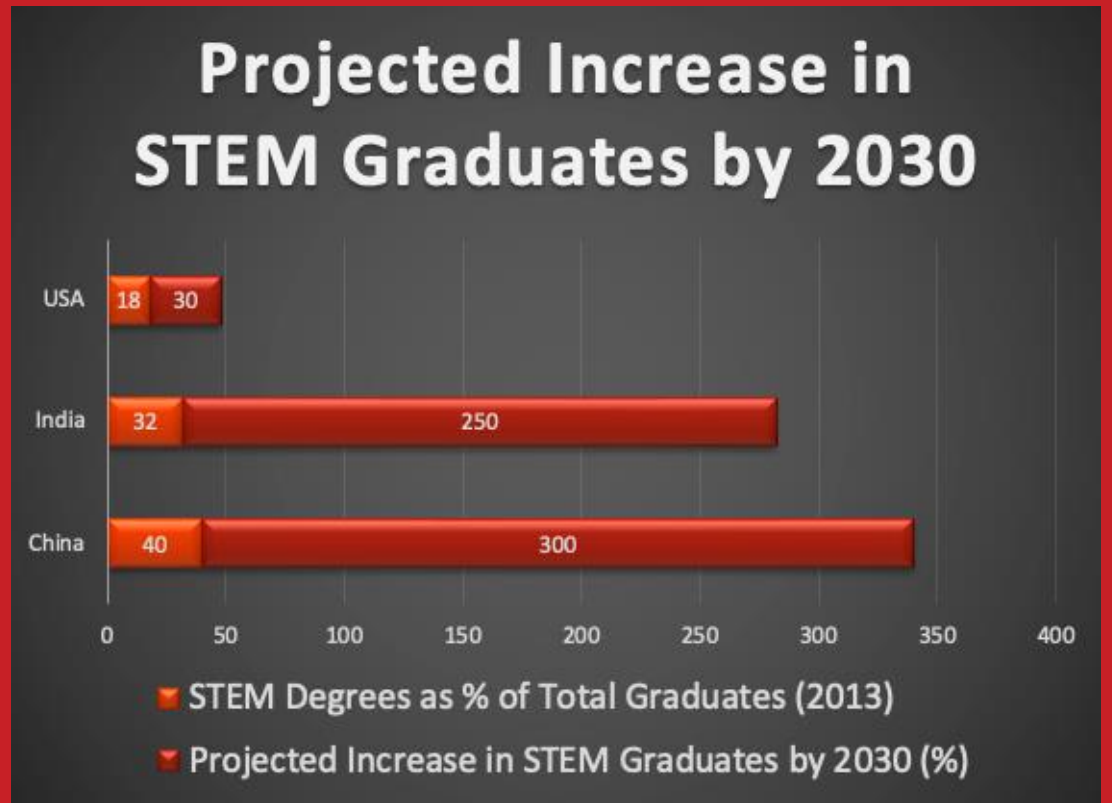


WHY NOW?

- **170M new tech roles by 2030 → AI/ML/data skills**
- **STEM grads: China +300% vs U.S. +30% (by 2030)**
- **KY STEM Educators: – 44% since 2010**

Sources: WEF; CSET; Bellarmine.





CADA Mission — Empower • Bridge • Transform

We equip students and faculty with practical AI & data analytics to close the talent gap and deliver lasting community impact.

1. **Empower:** Hands-on training + toolkits
2. **Bridge:** Job-ready skills & credentials
3. **Transform:** Classroom impact in KY

1. Advance Student Research Excellence

At WKU

8 Students Mentored for the WKU Student Scholar Showcase — April 5, 2025

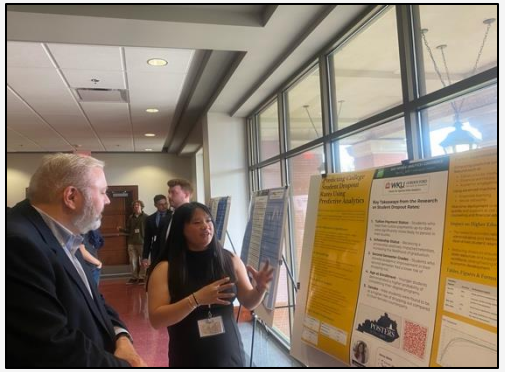
State Level (Kentucky)

2 Students Present at Posters at the Capitol — March 6, 2025

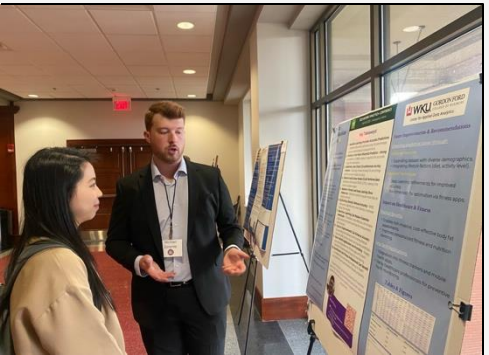
National

5 Students at the INFORMS Analytics Conference — April 6–8, 2025

At WKU



1. Parker Wiggins
2. Ryan Spychalski
3. John Costa
4. Cole Kidd
5. Hayden Coles
6. Christian Sander
7. Michael Delaney
8. Jenna Wells

Why is Sleeping Difficult?

Presenter: Parker Wiggins, Business Data Analytics

Advisor: Dr. Lily Popova Zhuhadar, Director of WKU Center for Applied Data Analytics

Research Objective & Significance

This study analyzes sleep quality by evaluating 13 factors that influence sleep patterns. The goal is to identify key contributors to poor sleep and suggest practical solutions. Findings benefit individuals, healthcare providers, and researchers studying sleep health.

Data & Methodology

- Dataset: 373 observations on sleep-related variables.
- Features: Sleep duration, age, BMI, stress levels, occupation, and health conditions.
- Target Variable: Sleep quality.
- Modeling Approach: Machine learning models including Random Forest, Naïve Bayes, Deep Learning, and Gradient Boost Tree.

Key Findings & Model Performance

Most Influential Factors:

1. Stress Levels: Strongest negative impact on sleep quality.
2. Sleep Duration: Those sleeping at least 7.95 hours report better sleep quality.
3. Age: Older individuals tend to sleep better.

Model Performance Metrics:

- Best Model: Random Forest with highest accuracy.
- Naïve Bayes performed well (93% accuracy).
- Gradient Boost Tree had the lowest RMSE, indicating strong predictive ability.

2025 INFORMS ANALYTICS+ CONFERENCE
April 6–8, 2025 | Indianapolis, Indiana

WKU GORDON FORD COLLEGE OF BUSINESS
Center for Applied Data Analytics

Key Takeaways!

1. **Stress is the Biggest Factor** – Higher stress leads to poorer sleep quality.
2. **Longer Sleep Improves Quality** – At least 7.95 hours of sleep leads to better rest.
3. **Age Positively Affects Sleep** – Older individuals report improved sleep.
4. **BMI and Occupation Play a Role** – Maintaining a healthy BMI and managing workload impact sleep.
5. **Random Forest Model is the Best Predictor** – Achieved the highest accuracy in sleep quality prediction.
6. **Naïve Bayes Model Performed Well** – Reached 93% accuracy, competing with advanced models.
7. **Gradient Boost Tree Had Lowest RMSE** – Provided the most precise predictions.
8. **Screen Time May Affect Sleep** – Further study is needed to confirm its impact.
9. **Sleep Disorders Need More Analysis** – Insomnia and other conditions could be key factors.
10. **AI Can Improve Sleep Health** – Predictive models can be used for personalized sleep solutions.

These findings highlight the importance of stress management, healthy habits, and AI-driven insights for better sleep quality.

Parker Wiggins
Email: parker.wiggins541@topper.wku.edu
LinkedIn: https://www.linkedin.com/in/pwigginsnovels/

Future Improvements & Recommendations

Enhancing predictive power through:

Additional Features:

- Incorporating sleep disorder data (insomnia, restless leg syndrome).
- Tracking screen time before sleep to analyze its impact.

Advanced Techniques:

- Deep Learning refinements for better accuracy.
- AI-based personalized sleep improvement recommendations.

Impact on Healthcare & Wellness

Practical Benefits:

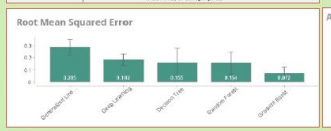
- Helps individuals understand and improve sleep quality.
- Provides healthcare professionals with data-driven insights.

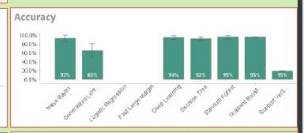
Long-Term Applications:

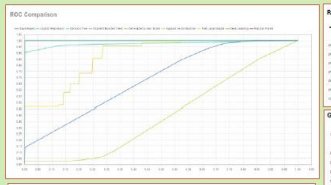
- Integration into wearable sleep trackers and health apps.
- Development of AI-powered sleep monitoring and intervention strategies.

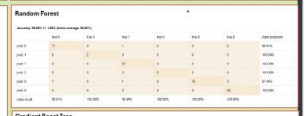
Tables & Figures

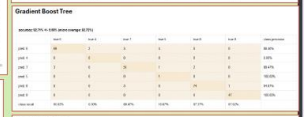
NAME	TYPE	DESCRIPTION
PERSON ID	Numerical	Unique identifier for each participant
GENDER	String	Sex of the patient (Male and Female)
AGE	Numerical	Age of the participant in years
OCCUPATION	String	Job of the participant
SLEEP DURATION	Numerical	Daily sleep duration of the participant
QUALITY OF SLEEP	Numerical	A subjective ranking of sleep quality (scaled 1-10)
PHYSICAL ACTIVITY	Numerical	Daily duration of physical activity for each participant measured in minutes
STRESS LEVEL	Numerical	A subjective ranking of the participant's stress (scaled 1-10)
BMI CATEGORY	String	The participant's BMI classification (Underweight, Normal, Overweight)
BLOOD PRESSURE	Numerical	The participant's blood pressure
HEART RATE	Numerical	The participant's resting heart rate (beats per minute)
DAILY STEPS	Numerical	The number of steps taken each day
SLEEP DISORDER	String	Whether the participant has a sleep disorder (None, Insomnia, or Sleep Apnea

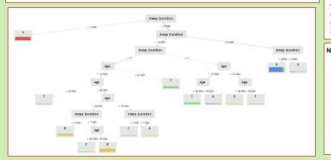












Social Media and Mental Health

A Data-Driven Approach to Improving Customer Experience

Presenter: *Ryan Spychalski, Finance & Business Data Analytics*

Advisor: *Dr. Lily Popova Zhuhadar, Director of WKU Center for Applied Data Analytics*

Research Objective & Significance

This study examines the impact of social media on mental health using predictive analytics. By leveraging machine learning models, it identifies key risk factors and trends, providing insights for businesses, policymakers, and healthcare professionals to mitigate negative effects while maximizing the potential for early intervention.

Data & Methodology

- Dataset: Publicly available Kaggle dataset focusing on university students.
- Classification Task: Identify mental health risk factors linked to social media use.
- Modeling Approach: Machine learning techniques including Naïve Bayes, Logistic Regression, Deep Learning, Decision Tree, and Random Forest.

Key Findings & Model Performance

Most Influential Factors:

1. Difficulty Concentrating: Strong correlation with excessive social media use.
2. Anxiety & Depression: Higher in frequent social media users.
3. Self-Comparison: More platforms used correlates with lower self-esteem.
4. Restlessness: Increased usage leads to higher stress levels.
5. Validation-Seeking Behavior: Minimal impact on overall mental health outcomes.

Model Performance Metrics:

- Best Accuracy: Random Forest (82.4%)
- Best Precision: Decision Tree (87.9%)
- Best Recall/Sensitivity: Deep Learning (85.7%)

2025 INFORMS ANALYTICS+ CONFERENCE
April 6–8, 2025 | Indianapolis, Indiana

WKU GORDON FORD COLLEGE OF BUSINESS
Center for Applied Data Analytics

Key Takeaways!

1. **Frequent Social Media Use Increases Mental Health Risks** – Higher usage links to anxiety, depression, and restlessness.
2. **Difficulty Concentrating is a Key Issue** – Social media users struggle with focus and attention.
3. **Self-Comparison Lowers Self-Esteem** – Comparing to others online leads to emotional distress.
4. **More Platforms, More Stress** – Using multiple social media apps increases mood fluctuations.
5. **Validation-Seeking Has Little Impact** – Likes and comments do not strongly predict mental health.
6. **AI Can Identify At-Risk Users** – Machine learning models predict mental health risks with high accuracy.
7. **Data Insights Enable Early Intervention** – Businesses and healthcare providers can offer targeted support.
8. **AI-Powered Tools Can Improve Mental Health** – Predictive analytics can support real-time monitoring.
9. **Ethical AI and Data Privacy Are Essential** – Responsible tech use must protect user well-being.
10. **Organizations Must Act** – Businesses and policymakers should design healthier digital spaces.

These findings highlight the need for responsible social media practices and data-driven mental health strategies to create safer, more supportive digital experiences.

Ryan Spychalski

- Email: ryanspychalski95@gmail.com
- Phone: (931) 446-3214
- LinkedIn: linkedin.com/in/ryan-spychalski-606264223

Future Improvements & Recommendations

Enhancing predictive power through:

Additional Features:
- Inclusion of demographic and behavioral data.
- Real-time engagement tracking and sentiment analysis.

Advanced Techniques:
- Deep Learning refinements for better accuracy.
- AI-driven mental health screening tools.

Impact on Society & Policy

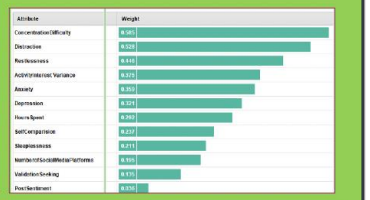
Business & Public Health Implications:
- Organizations can leverage data to develop mental health programs.
- Healthcare providers can identify high-risk individuals early.

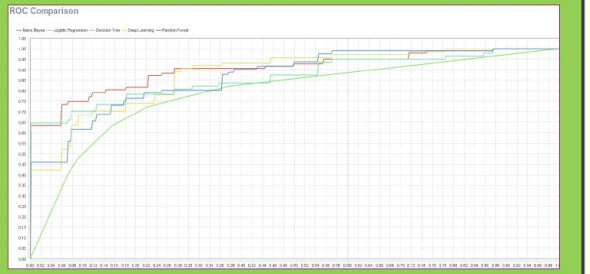
Policy Considerations:
- Ethical AI regulations for user data protection.
- Integrating predictive analytics into mental health frameworks.

Tables & Figures

Variable Set

Attribute	Description	Data Type
Concentration Difficulty	Scaled rating 1-5 of subjective view that it is difficult to concentrate	Ordinal
Distraction	Scaled rating 1-5 of subjective view that one is distracted easily	Ordinal
Restlessness	Scaled rating 1-5 of subjective view that one is restless when not using social media	Ordinal
Activity Interest Variance	Scaled rating 1-5 of subjective view that activity interests fluctuate	Ordinal
Anxiety	Scaled rating 1-5 of subjective distress experienced by worries	Ordinal
Depression	Scaled rating 1-5 of subjective view that one is feeling down or depressed often	Ordinal
Hours Spent	Categorical estimate of hours spent on social media: <2 hours, 2-3 hours, and >5 hours	Polynomial
Sleeplessness	Scaled rating 1-5 of subjective view that one is facing problems with sleep often	Ordinal
Self Comparison	Scaled rating 1-5 of subjective view that one is comparing self to others	Ordinal
Validation Seeking	Scaled rating 1-5 of subjective view that one is seeking validation of self from others	Ordinal
Post Sentiment	Scaled rating 1-5 of subjective view of how one feels about posts on social media that seek validation	Ordinal
Frequency	Report of how often a person uses social media on a scale of 1-5	Ordinal





Model	Accuracy	Classification Error	AUC	Precision	Recall	F-Measure	Sensitivity (TPR)	Specificity (TNR)	False Positives	False Negatives
Naive Bayes	81.6%	18.4%	86.7%	83%	83%	82.9%	82.9%	79.4%	13	12
Logistic Regression	81.7%	18.3%	86.6%	85%	84.9%	84.3%	84.9%	76.9%	12	13
Deep Learning	78.1%	21.9%	87.9%	76.5%	85.7%	79.7%	85.7%	71.7%	11	19
Decision Tree	79.2%	20.8%	N/A	87.9%	76.6%	81.5%	76.6%	81.7%	9	19
Random Forest	82.4%	17.6%	90.3%	85.2%	84.9%	84.4%	84.9%	81.2%	11	13

Predicting hazardous near-Earth objects using machine learning for planetary defense

Presenter: ***John Costa***, *Business Data Analytics*
Advisor: *Dr. Lily Popova Zhuhadar, Director of WKU Center for Applied Data Analytics*

2025 INFORMS ANALYTICS+ CONFERENCE
April 6–8, 2025 | Indianapolis, Indiana

WKU GORDON FORD COLLEGE OF BUSINESS
Center for Applied Data Analytics

Research Objective & Significance

- The study develops a machine learning model to classify Near-Earth Objects (NEOs) as hazardous or non-hazardous based on their physical and orbital characteristics.
- NEOs, including asteroids and comets, often pass close to Earth, and while most pose no threat, some could have catastrophic impacts.
- This research provides data-driven insights for planetary defense agencies, helping prioritize monitoring and mitigation efforts for high-risk objects.
- It is relevant for state legislators, emphasizing the role of machine learning in planetary defense strategies and funding allocation for science and technology programs.

Data & Methodology

- The study uses NASA's certified NEO dataset, containing 90,836 unique observations of asteroids and comets.
- Features analyzed include velocity, diameter, and proximity to Earth, which are key indicators of potential hazard.

Machine learning models used:
- Decision Trees
- Random Forests (chosen as the best model).

Key Findings & Model Performance

Random Forest Model Performance:

- **Recall**: 40.50% (improved ability to detect hazardous NEOs)
- **Precision**: 67.29% (balanced accuracy in hazardous classifications)
- **AUC Score**: 0.941 (excellent ability to distinguish hazardous vs. non-hazardous objects)

Key Features Influencing Predictions:
- Proximity to Earth (closer objects pose a higher risk)
- Velocity (higher speeds increase potential impact damage)
- Diameter (larger objects pose greater threats)

Correlation Analysis:
- Strong Correlations: Between minimum and maximum estimated diameters
- Moderate Correlations: Absolute magnitude and diameter (negative correlation)

Key Takeaways!

1- Deployment & Practical Applications:

- The Random Forest model was integrated with NASA's CNEOS API for real-time hazard detection.
- Hosted on a cloud-based platform for scalability and accessibility.
- A user-friendly web dashboard displays predictions, feature importance, and hazard scores for decision-makers.
- Continuous monitoring and updates ensure long-term reliability.

2- Impact on Planetary Defense & Policy Recommendations

- This research offers a critical tool for planetary defense teams to prioritize hazardous NEO monitoring.
- It provides legislators and policymakers with a cost-effective, data-driven approach to space safety and funding decisions.
- By leveraging AI in planetary defense, states can contribute to global efforts in mitigating space-related risks.

POSTERS At-The-Capitol

John Costa
Email: johncosta97@gmail.com
LinkedIn: linkedin.com/in/john-costa-8b575b230
Phone: (270) 799-3433

Model Evaluation & Enhancements

Decision Tree Model Issues:

- It performed well in predicting non-hazardous NEOs but struggled with recall for hazardous cases.

Random Forest Model Refinements:

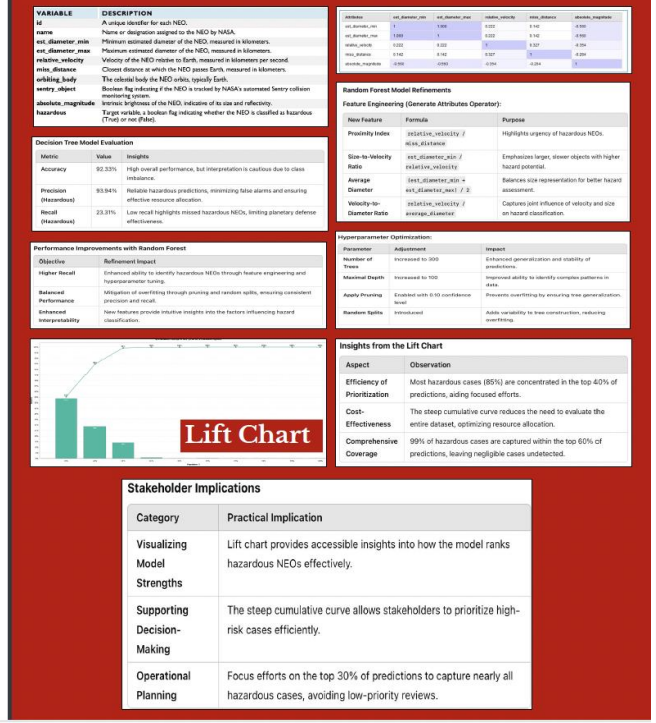
- Feature engineering was applied (e.g., proximity index, velocity-to-diameter ratio).
- Hyperparameter tuning improved overall accuracy and recall.
- Class balancing helped reduce misclassifications.

Conclusion

This research demonstrates the power of machine learning in planetary defense by enabling real-time risk assessment of Near-Earth Objects.

By continuously improving prediction models and integrating real-time data, decision-makers can enhance planetary safety efforts, allocate resources effectively, and ensure proactive hazard mitigation strategies.

Tables, Figures & Formulas



Enhancing Airline Passenger Satisfaction Using Predictive Analytics

A Data-Driven Approach to Improving Customer Experience

*Presenter: **Cole Kidd**, Business Data Analytics*

Advisor: Dr. Lily Popova Zhuhadar, Director of WKU Center for Applied Data Analytics, Western Kentucky University

Research Objective & Significance

This study develops a predictive model using Gradient Boosted Trees to assess airline passenger satisfaction based on over 100,000 survey responses. The model identifies key satisfaction drivers, such as inflight entertainment and class, with an accuracy of 95.7%.

Airlines can leverage these insights to improve customer retention, service quality, and profitability.

Data & Methodology

- **Dataset**: 103,903 airline passenger surveys (Kaggle)
- **Classification Task**: Predict passenger satisfaction levels
- **Modeling Approach**: Machine learning using Gradient Boosted Trees

Key Findings & Model Performance

Most Influential Factors:

- **Class Type**: First-class passengers report higher satisfaction.
- **Inflight Entertainment**: Strongly impacts passenger experience.
- **Food & Drink**: Affects overall satisfaction significantly.
- **Flight Distance**: Long-haul flights show higher satisfaction.
- **On-Time Performance**: Delays negatively impact satisfaction.

Model Performance Metrics:

- **Accuracy**: 95.7%
- **Precision**: 95.8%
- **Recall** (Sensitivity): 94.4%
- **Specificity**: 96.8%
- **Classification Error**: 4.3%.

2025 INFORMS ANALYTICS+ CONFERENCE
April 6–8, 2025 | Indianapolis, Indiana

WKU GORDON FORD COLLEGE OF BUSINESS
Center for Applied Data Analytics

Key Takeaways from this Research!

1. **Class Type Matters** – First-class passengers report higher satisfaction than economy travelers, highlighting the value of premium services.
2. **Inflight Entertainment is Crucial** – Quality entertainment significantly boosts passenger satisfaction, urging airlines to invest in engaging content.
3. **Food & Beverage Service Enhances Experience** – Better meals and drinks improve satisfaction, emphasizing the need for high-quality catering.
4. **Longer Flights Correlate with Higher Satisfaction** – Long-haul passengers report greater satisfaction, likely due to better amenities.
5. **Punctuality Impacts Perception** – Flight delays hurt satisfaction, stressing the importance of operational efficiency.
6. **Loyal Customers Show Higher Retention Rates** – Frequent flyers exhibit higher satisfaction, reinforcing the need for strong loyalty programs.
7. **Gradient Boosted Trees Model Effectiveness** – The model achieved 95.7% accuracy, making it a powerful tool for improving airline service.
8. **Data-Driven Insights Can Enhance Airline Strategies** – Predictive analytics help airlines refine marketing, improve service, and boost retention.
9. **Potential for Industry-Wide Adoption** – AI-driven analytics can transform customer experience across various service industries.
10. **Policy Implications for State Legislators** – Data-driven decisions can improve Kentucky's airline services and economic growth.

These insights guide airlines in enhancing customer experience, boosting loyalty, and increasing profitability through data-driven improvements.

Cole Kidd

- Email: colekidd923@gmail.com
- Phone: (270) 217-5626
- LinkedIn: https://www.linkedin.com/in/colekidd

Future Improvements & Recommendations

Enhancing predictive power through:

Additional Features:

-Seat comfort, loyalty status, and customer feedback analysis
- Real-time sentiment tracking from social media reviews

Advanced Techniques:

- Ensemble Learning & Neural Networks for better prediction
-AI-Driven Personalization for tailored passenger experiences

Impact on the Airline Industry & Policy

Strategic Benefits for Airlines:

- Better allocation of resources to improve service aspects
- Increased revenue by enhancing customer retention
- Data-driven marketing strategies for targeted advertising

Implications for Policymakers:

- Supports state-wide economic growth by improving air travel experiences
- Sets a precedent for using predictive analytics in other sectors

Conclusion

The Gradient Boosted Trees model (94.4% accuracy) helps airlines enhance key satisfaction drivers, boosting retention and profits. Future improvements include more predictors and updated data to adapt to changing preferences.

Tables & Figures

Smart Workforce Management: Leveraging AI for Employee Retention

Presenter: Hyden Coles, Business Data Analytics

Advisor: Dr. Lily Popova Zhuhadar, Director of WKU Center for Applied Data Analytics

Research Objective & Significance

This study explores employee attrition through the application of machine learning models, leveraging predictive analytics to identify employees at risk of leaving. By analyzing key factors such as job satisfaction, work-life balance, tenure, and overtime, businesses can implement proactive retention strategies to reduce turnover costs and enhance workforce stability. The findings provide actionable insights for HR professionals, business leaders, and policymakers, enabling data-driven decision-making to improve employee engagement and organizational resilience..

Data & Methodology

- Dataset: 74,498 employee records.
- Key Features: Work-life balance, job satisfaction, tenure, salary, overtime, performance rating.
- Modeling Approach: Auto-Modeling with Altair AI Studio.
- Best Model: Gradient Boosted Trees with 66.59% accuracy.

Key Findings & Model Performance

Most Influential Factors:

- High-risk attrition factors: Low job satisfaction, poor work-life balance, frequent overtime.
- Employees with shorter tenure more likely to leave.

Model Performance Metrics:

- Accuracy: 66.59%
- Precision (Stayed): 80.81%
- Recall (Left): 89.36%

Model effectively identifies employees at risk of leaving.

WKU GORDON FORD COLLEGE OF BUSINESS
Center for Applied Data Analytics

Key Takeaways!

1. **Shorter Tenure Increases Risk** – Employees with less experience are more likely to leave.
2. **Low Job Satisfaction Leads to Turnover** – Dissatisfied employees are at higher risk of leaving.
3. **Frequent Overtime Increases Attrition** – Employees working excessive overtime are more likely to quit.
4. **AI Helps Predict Attrition** – Machine learning models can proactively identify at-risk employees.
5. **HR-Driven Data Strategies Work** – Data-driven retention strategies improve workforce stability.
6. **High Salaries Don't Guarantee Retention** – Compensation alone does not ensure employee loyalty.
7. **Exit Interviews Provide Critical Insights** – Understanding why employees leave enhances prediction models.
8. **Work-Life Balance Matters** – Poor work-life balance significantly contributes to turnover.
9. **Real-Time Monitoring is a Game Changer** – AI-powered dashboards help HR teams make proactive decisions.
10. **Reducing Turnover Saves Costs** – Replacing an employee can cost 20–30% of their annual salary.

These findings reinforce the importance of predictive analytics in workforce stability and retention.

Hyden Coles

Email: hayden.coles986@topper.wku.edu
LinkedIn: https://www.linkedin.com/in/hayden-coles-5840ba256/

Future Improvements & Recommendations

Enhancing predictive power with:

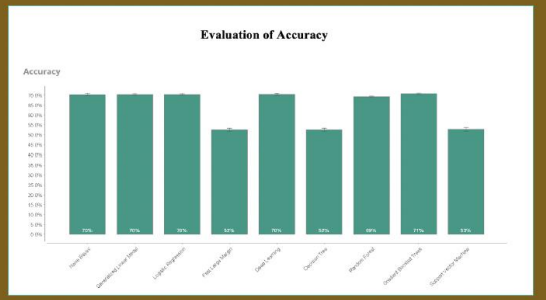
Additional Features:

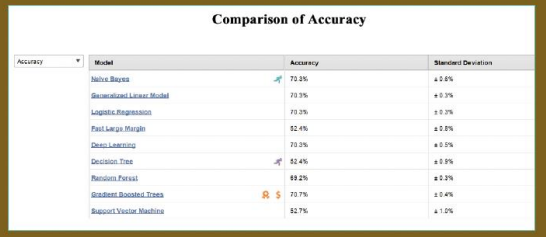
- Employee engagement, training, recognition.
- AI-powered dashboards for real-time attrition monitoring.
- Exit interview insights for better training models.

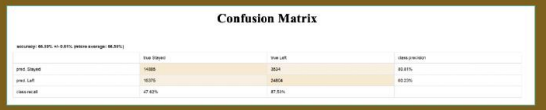
Impact on Workforce & Business Stability

- Predictive modeling supports workforce stability and reduces turnover costs.
- Helps HR teams proactively address high-risk employees.
- Saves businesses significant costs on hiring and training new employees.
- Supports organizational resilience through data-driven decision-making.

Tables & Figures







Leveraging Machine Learning to Enhance Workforce Stability and Retention

Presenter: Christian Sander, Business Data Analytics

Advisor: Dr. Lily Popova Zhuhadar, Director of WKU Center for Applied Data Analytics

Research Objective & Significance

This study examines employee attrition using machine learning models. By predicting workforce turnover, businesses can proactively address risk factors and improve employee retention strategies. Findings support HR professionals, business leaders, and policymakers in workforce planning.

Data & Methodology

- Dataset: 4,653 employee records from India.
- Features: Gender, education, payment tier, tenure, and hiring date.
- Target Variable: Employee attrition within two years.
- Modeling Approach: Logistic Regression, Random Forest, and Deep Learning.
- Best Model: Logistic Regression with 85.62% recall.

Key Findings & Model Performance

Most Influential Factors:

1. Employees hired after 2017 are more likely to leave.
2. Higher payment tier employees have higher attrition rates.
3. Employees with shorter tenure are at greater risk.

Model Performance Metrics:

- Best Model: Logistic Regression (85.62% recall).
- Decision Tree and Random Forest performed well but were less effective.
- Further optimization may improve prediction accuracy.

WKU GORDON FORD COLLEGE OF BUSINESS
Center for Applied Data Analytics

Key Takeaways!

1. **Newer Hires Leave More** – Employees hired after 2017 are more likely to quit.
2. **Higher Payment Tiers See More Attrition** – Well-paid employees leave at higher rates.
3. **Short Tenure Increases Risk** – Employees with less experience are more likely to leave.
4. **Logistic Regression is the Best Model** – Achieved 85.62% recall, outperforming others.
5. **Random Forest and Decision Tree Performed Well** – But did not surpass Logistic Regression.
6. **Turnover is Costly** – Replacing employees costs 20–30% of their annual salary.
7. **HR Can Use AI to Retain Employees** – Predictive models help prevent unwanted attrition.
8. **Employee Engagement is Crucial** – Work anniversaries and training improve retention.
9. **Exit Interviews Provide Key Insights** – Understanding why employees leave can refine predictive Models.
10. **AI-Powered Dashboards Can Help** – Real-time monitoring can support HR strategies.

These findings highlight the importance of predictive analytics in workforce stability and retention.

Christian Sander
Gmail: Chris.sander8@gmail.com
Cell Phone: 502-500-0481
LinkedIn: https://www.linkedin.com/in/christiansander02

Future Improvements & Recommendations

Enhancing predictive power through:

Additional Features:

- Employee engagement, work anniversaries, training, and recognition.
- Exit interview insights for better model training.

Advanced Techniques:

- Cross-validation and hyperparameter tuning for improved accuracy.
- AI-powered dashboards for real-time attrition monitoring.

Impact on Workforce & Business Stability

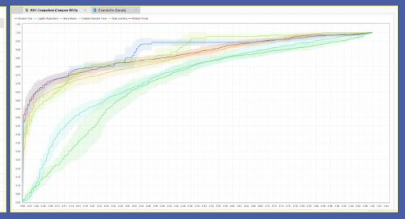
Practical Benefits:

- Helps HR teams proactively address turnover risks.
- Saves businesses significant costs in hiring and training new employees.

Long-Term Applications:

- Supports workforce stability and organizational resilience.
- Provides policymakers with insights for job retention strategies.

Tables & Figures



Metric	Details
Best Model	Logistic Regression using Sample Operator with 70% training and 30% testing data split.
Performance Metrics	- Recall (True Positives): 85.62% (highest recall achieved). - Accuracy: 67.71%.
Prediction Insights	- Predicted 411 out of 480 employees who left (true positives). - Missed predicting 69 employees who left.
Key Predictors	- Employees are likely to leave if: - Joining year > 2017.5 (2018). - Payment Tier > 1.5 (Tier 2 or 3).
AUC and ROC	- AUC: 0.734 (decent performance). - ROC indicates room for improvement (does not hug the Y-axis).
Business Impact	- Identifies 85.62% of at-risk employees, potentially saving the company significant costs.
Model Selection Rationale	- Recall (85.62%) prioritized over AUC and ROC for HR-focused decision-making.

Predicting Body Fat Percentage Using Machine Learning Models in AI Studio™

Presenter: *Michael Delaney, Business Data Analytics*

Advisor: *Dr. Lily Popova Zhuhadar, Director of WKU Center for Applied Data Analytics*

Research Objective & Significance

This study applies machine learning to predict body fat percentage using physical measurements. The goal is to develop a non-invasive and cost-effective alternative to traditional body fat assessment methods. Findings benefit healthcare providers, fitness professionals, and individuals seeking accessible health insights.

Data & Methodology

- Dataset: Kaggle's Body Fat Prediction dataset (252 male subjects).
- Features: Age, weight, height, and body circumferences (neck, chest, abdomen, hip, thigh, knee, ankle, biceps, forearm, wrist).
- Target Variable: Body fat percentage.
- Modeling Approach: Machine learning models including Generalized Linear Model (GLM), Random Forest, Deep Learning, and Decision Tree.

Key Findings & Model Performance

Most Influential Factors:

1. Density: Strong negative correlation (-0.988) with body fat.
2. Abdomen Circumference: Strong positive correlation (0.813) with body fat.
3. Chest Circumference: Significant predictor of body fat percentage.

Model Performance Metrics:

Best Model: Generalized Linear Model (GLM) with lowest RMSE (0.375).

Random Forest & Deep Learning also performed well when Density was included.

Excluding Density significantly reduced model accuracy.

2025 INFORMS ANALYTICS+ CONFERENCE
April 6–8, 2025 | Indianapolis, Indiana

Key Takeaways!

1. **Machine Learning Provides Accurate Predictions** – Models effectively estimate body fat percentage using simple physical measurements.
2. **Density is the Most Influential Predictor – Strong** correlation (-0.988) makes it essential for accurate body fat estimation.
3. **Abdomen and Chest Circumferences Are Key Factors** – Strongly linked to body fat percentage and improve model accuracy.
4. **Generalized Linear Model (GLM) Performs Best** – Achieves the lowest RMSE (0.375), making it the most reliable model.
5. **Random Forest and Deep Learning Show Potential** – Perform well but require more data for improvement.
6. **Excluding Density Reduces Accuracy** – RMSE increases significantly when Density is removed from the model.
7. **Machine Learning Can Replace Expensive Methods** – Provides a cost-effective, non-invasive alternative to traditional body fat measurement techniques.
8. **Future Improvements Can Enhance Accuracy** – Expanding the dataset and including lifestyle factors (diet, activity level) can improve predictions.
9. **Real-World Applications Are Promising** – Models can be integrated into fitness apps and health trackers for personalized body fat monitoring.
10. **Healthcare and Fitness Professionals Can Benefit** – Enables improved decision-making in health assessments and personalized fitness programs.

These findings highlight the potential of machine learning in providing accessible and accurate body fat assessments without the need for expensive medical equipment.

Michael Delaney

Email: mikedelaney02@gmail.com
Cell: (270) 349-7189
LinkedIn: https://www.linkedin.com/in/michael-delaney-924437254/

WKU GORDON FORD COLLEGE OF BUSINESS
Center for Applied Data Analytics

Future Improvements & Recommendations

Enhancing predictive power through:

Additional Features:

- Expanding dataset with diverse demographics.
- Integrating lifestyle factors (diet, activity level).

Advanced Techniques:

- Deep Learning refinements for improved accuracy.
- Real-time body fat estimation via fitness apps.

Impact on Healthcare & Fitness

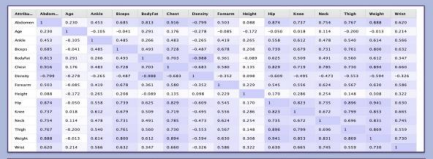
Practical Benefits:

- Enables non-invasive, cost-effective body fat assessments.
- Improves personalized fitness and nutrition planning.

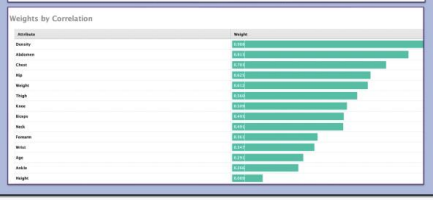
Long-Term Applications:

- Integration into fitness trackers and mobile health apps.
- Use by healthcare professionals for preventive health monitoring.

Tables & Figures



Model	With Density (RMSE)	Without Density (RMSE)
Generalized Linear Model	0.375	3.887
Deep Learning	1.903	4.398
Decision Tree	1.646	4.849
Random Forest	2.181	4.649



Predicting College Student Dropout Rates Using Predictive Analytics

Presenter: Jenna Wells, *Business Data Analytics & Accounting, Western Kentucky University*
Advisor: Dr. Lily Popova Zhuhadar, *Director of WKU Center for Applied Data Analytics, Western Kentucky University*

Research Objective & Significance

- The study develops a predictive model to forecast college student dropout rates using machine learning.
- By identifying at-risk students early, institutions can implement timely interventions to reduce dropout rates and increase graduation rates.
- The research is particularly relevant to Kentucky policymakers, as student retention directly impacts state economic growth.

Data & Methodology

- Dataset: 4,424 student records from Kaggle with 36 academic, demographic, and socioeconomic variables.
- Classification Task: Predict whether a student will graduate, drop out, or remain enrolled.
- Modeling Approach: Machine learning techniques in RapidMiner Auto Model, with Logistic Regression selected as the best-performing model.

Model Evaluation & Insights

- ROC Curve Analysis: No single model significantly outperformed others, but Logistic Regression provided a balanced trade-off between accuracy and interpretability.
- Confusion Matrix: The model was better at predicting graduates than dropouts, suggesting a need for improved dropout predictions.
- Lift Chart: Showed high-confidence predictions for graduates but lower accuracy in identifying dropouts, indicating the need for further refinement.

Key Findings & Model Performance

Model	Accuracy ↓
Logistic Regression	88.2%
Generalized Linear Model	88.2%
Deep Learning	86.5%

Model	Precision ↓
Gradient Boosted Trees	90.7%
Logistic Regression	90.7%
Generalized Linear Model	90.5%

Model	Sensitivity ↓
Generalized Linear Model	92.5%
Logistic Regression	92.3%
Deep Learning	91.1%

2025 INFORMS ANALYTICS+ CONFERENCE
April 6–8, 2025 | Indianapolis, Indiana

WKU GORDON FORD COLLEGE OF BUSINESS
Center for Applied Data Analytics

Key Takeaways from the Research on Student Dropout Rates:

1. **Tuition Payment Status** – Students who kept their tuition payments up-to-date were significantly more likely to persist in their studies.
2. **Scholarship Status** – Receiving a scholarship positively impacted retention, increasing the likelihood of graduation.
3. **Second-Semester Grades** – Students who showed academic improvement in their second semester had a lower risk of dropping out.
4. **Age at Enrollment** – Younger students demonstrated a higher probability of completing their degree programs.
5. **Gender** – Male students were found to be at a higher risk of dropping out compared to their female counterparts.

POSTERS At-The-Capitol

Jenna Wells

Jenna Wells
- Email: Jennawells14@icloud.com
- Phone: (859) 314-5753
- LinkedIn: linkedin.com/in/jennawells2025

Future Improvements & Recommendations

Enhancing predictive power by including additional features such as:

- Extracurricular involvement
- Financial aid history
- Academic engagement metrics

Using advanced machine learning techniques, such as:

- Ensemble learning
- Neural networks

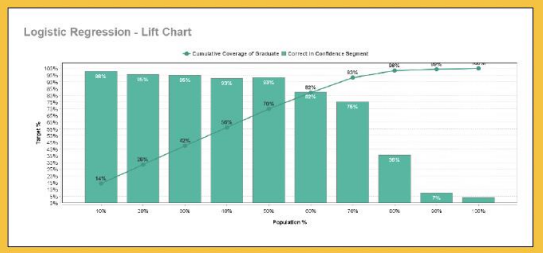
Real-time deployment to enable universities to act quickly and support at-risk students through academic counseling and financial assistance.

Impact on Higher Education & Policy

- The research provides valuable insights for university administrators and Kentucky legislators to create data-driven student retention policies.
- Reducing dropouts ensures efficient allocation of state resources and supports a more educated workforce, strengthening Kentucky's long-term economic development.

Tables, Figures & Formulas

Population Segment	Accuracy	Performance Insights
Top 10%	98%	Model performs exceptionally well with high-confidence predictions.
Top 50%	70%	Accuracy remains strong, indicating reliable predictions for a larger portion of the population.
80%-90%	36%	Performance significantly declines as confidence decreases.
90%-100%	7%	Very weak predictions in the lowest-confidence segment, highlighting the need for improvement.



State Level (Kentucky)

POSTERS At-The-Capitol

- GFCB Shines at the 2025 Posters at the Capitol!
  - Planetary Defense (NASA NEOs) — John Costa
  - Student Retention — Jenna Wells' predictors of dropout ⟶ policy actions

2025 INFORMS ANALYTICS+ CONFERENCE

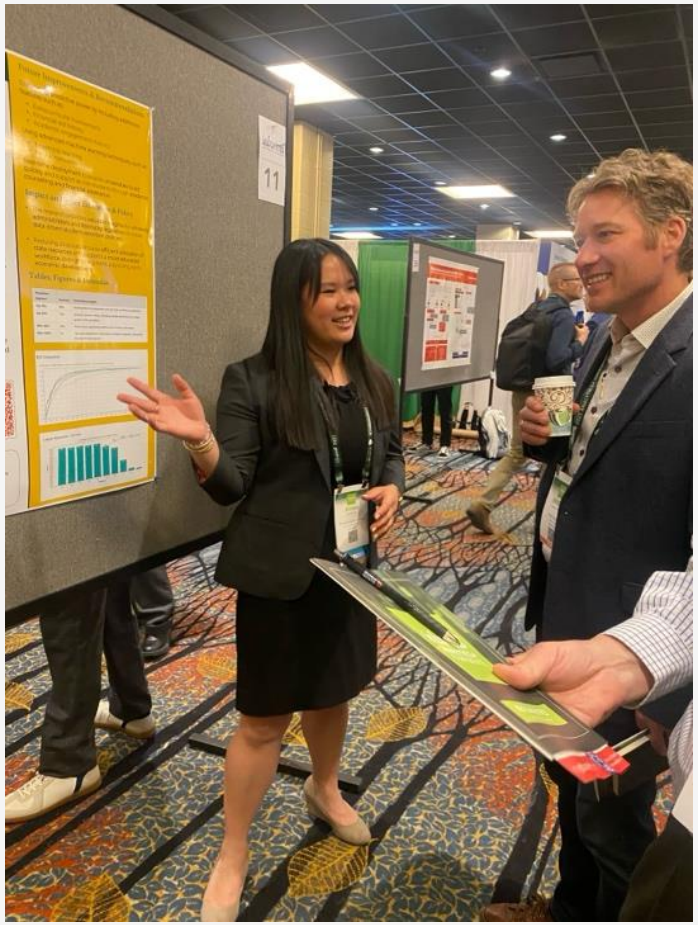
April 6–8, 2025 | Indianapolis, Indiana

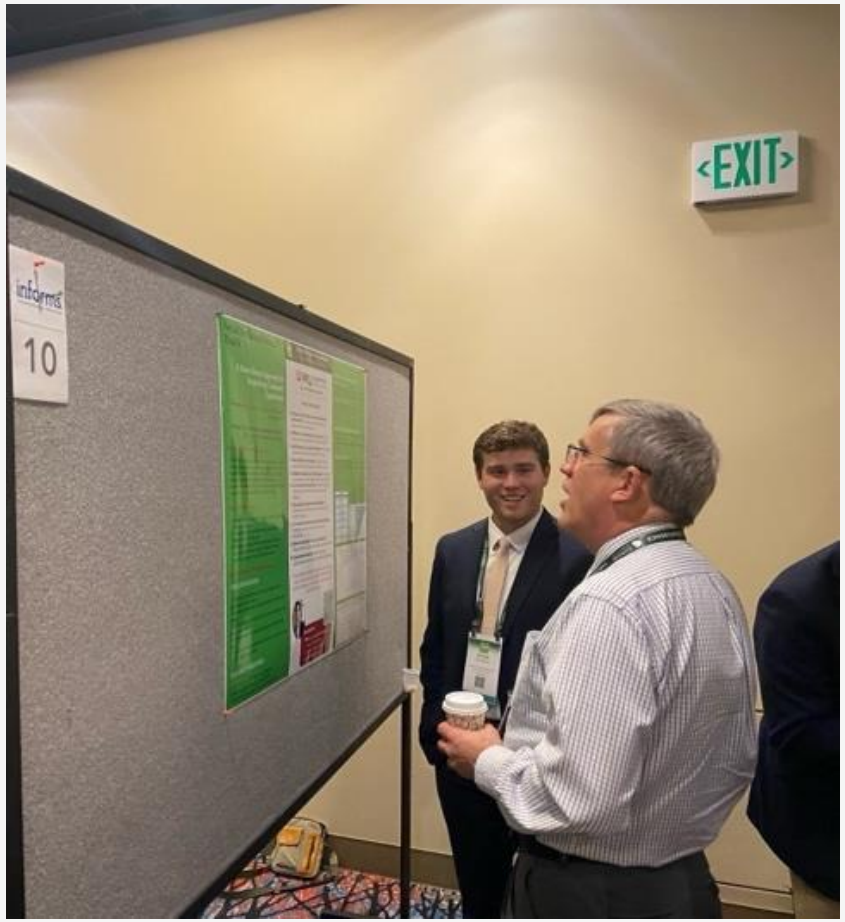
Jenna Wells, Ryan Spychalski, Michael Delany, John Costa, and Cole Kidd delivered research presentations on topics ranging from mental health and planetary defense to student retention, airline satisfaction, and healthcare analytics.

INFORMS President Dave Hunt (right) and Shaun Doheney, AWS Principal Analytics Leader and INFORMS Conference Chair (left)!

2025 INFORMS ANALYTICS+ CONFERENCE

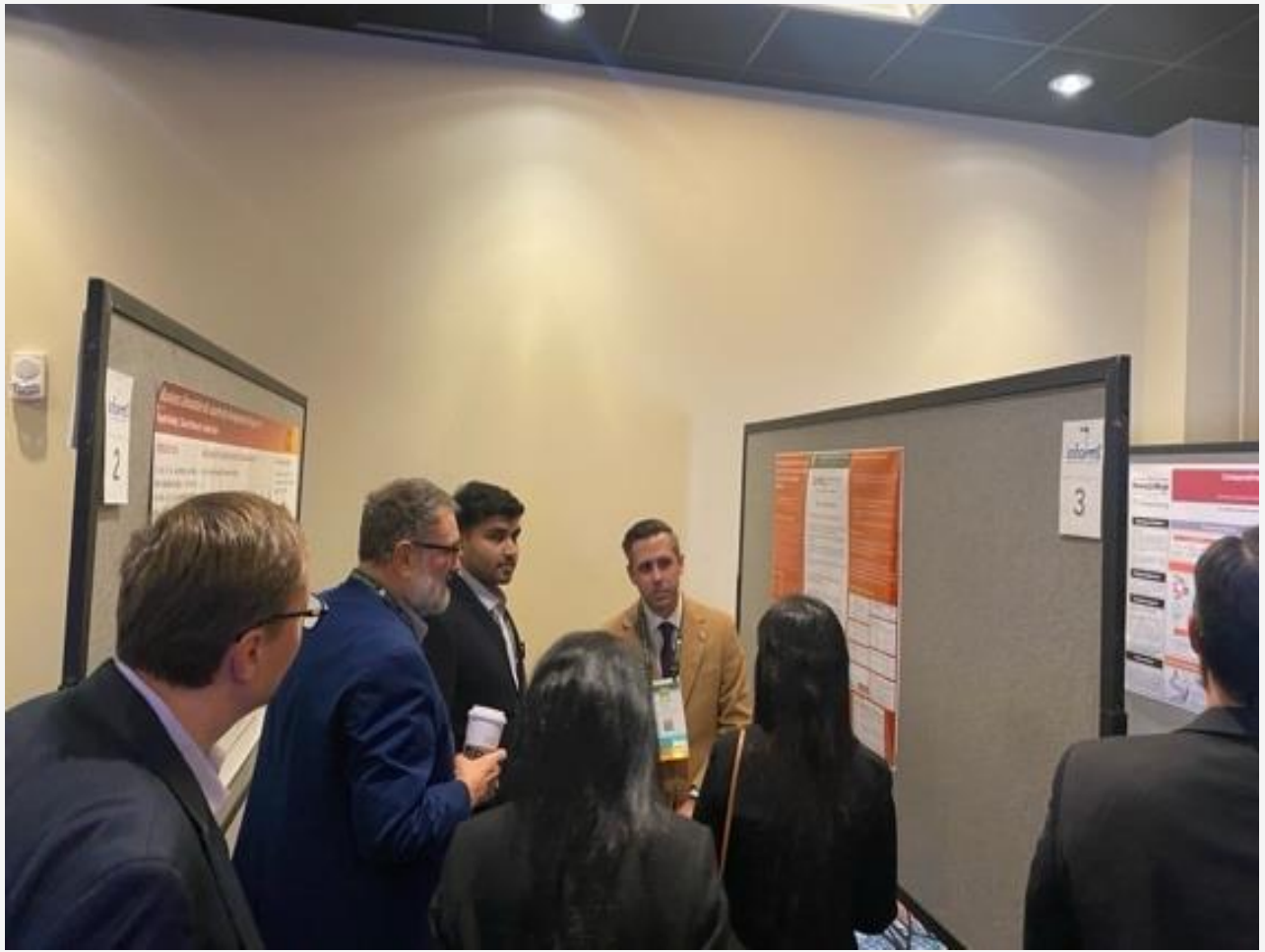
April 6–8, 2025 | Indianapolis, Indiana





2025 INFORMS ANALYTICS+ CONFERENCE

April 6–8, 2025 | Indianapolis, Indiana



- Michael Delaney
- Cole Kidd
- Jenna Wells
- John Costa
- Ryan Spychalsky

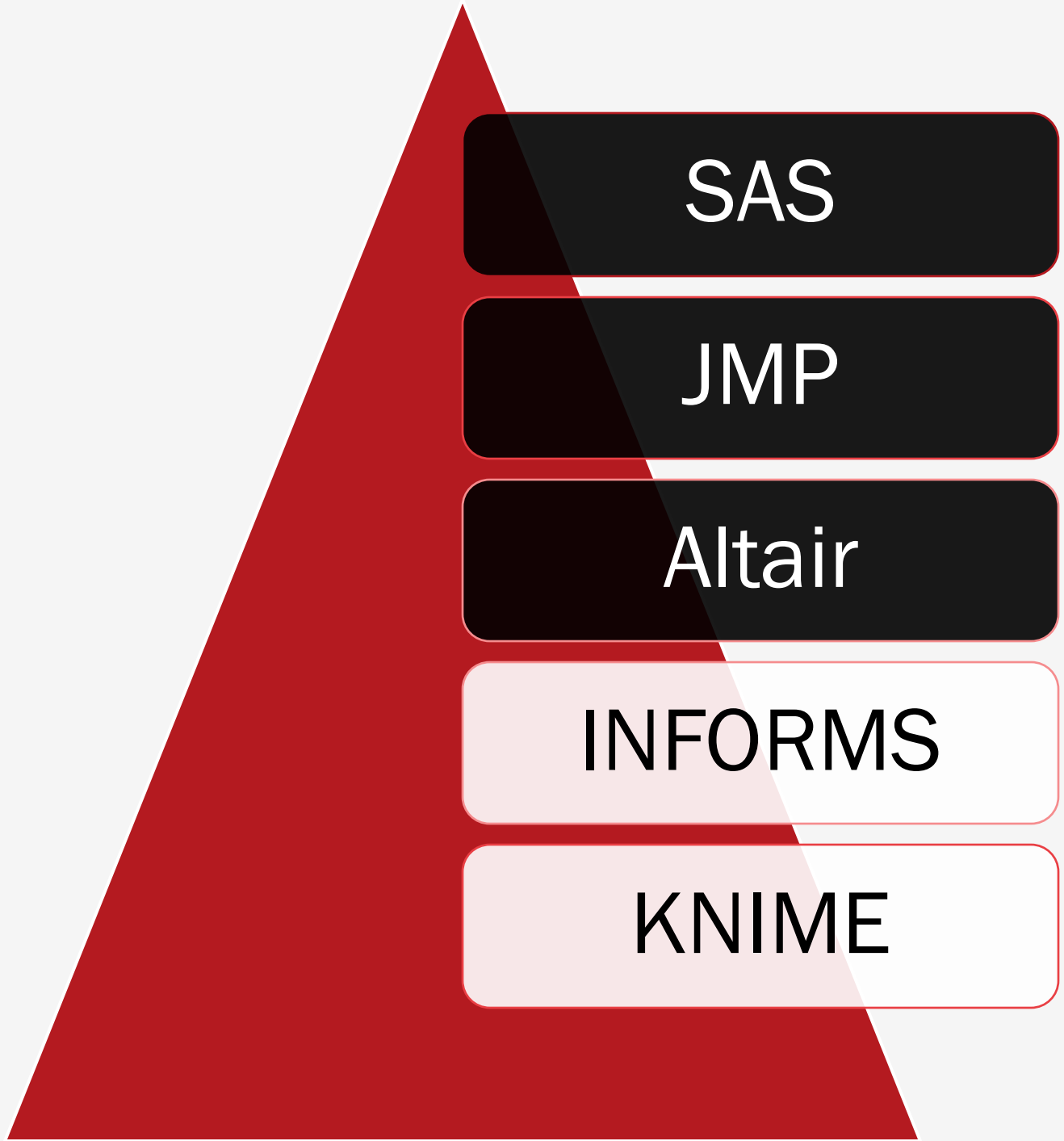
- Remember these names. We predict they will be the future leaders of analytics!

Board Ask

What’s one high-impact way your company could engage WKU students this year?

1. Paid internships
2. Micro-internships (2–4 weeks)

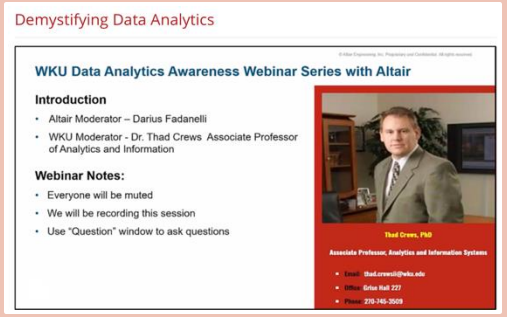
2. Forge High-Impact Partnerships

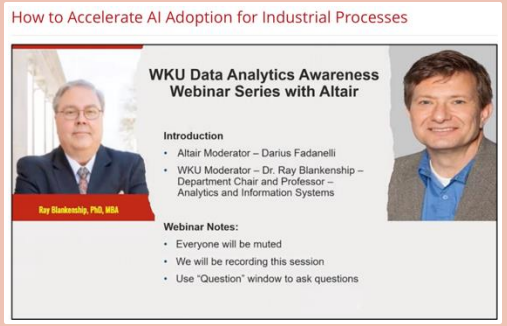


Delivered Analytics Training Sessions

2024 – WKU Data Analytics Awareness Webinar Series with Altair AI

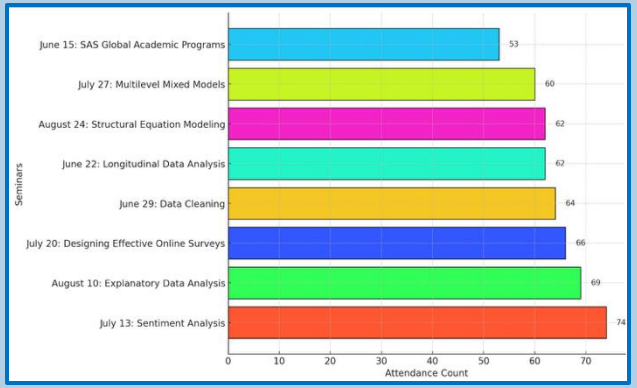
- Demystifying Data Analytics
- How to Accelerate AI Adoption for Industrial Processes
- Unlocking AI/ML Potential for Manufacturing





2023 – CADA Launched a Virtual Summer Enrichment Series for WKU faculty, staff, and GAs

- Enabling Data Science & Analytics - Lori Downen
- Longitudinal Data Analysis – Jackie Johnson
- Data Cleaning – Nancy Rausch
- Sentiment Analysis (Text Mining) – Tom Grant
- Designing Effective Online Surveys – Nancy Rausch
- Multilevel Mixed Models – Jackie Johnson
- Exploratory Data Analysis – Tom Grant
- Structural Equation Modeling – Cat Truxillo



- The chart highlights attendance for the 2024 seminars held from June to August.
- Sentiment Analysis recorded the highest attendance with 74 participants, while SAS Global Academic Programs had the lowest at 53.
- Other popular sessions included Explanatory Data Analysis (69 attendees) and Designing Effective Online Surveys (66 attendees), showcasing a strong interest in practical and applied data analytics skills.

2022 – 2nd Annual SAS Training Workshops

2nd Annual SAS Training Workshops: Delivered across multiple WKU colleges, this series expanded access to analytics training:

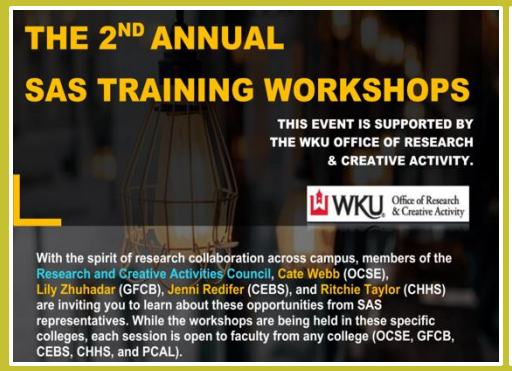
- GFCB: SAS Studio, Fraud Detection, R & Python in SAS
- CEBS: SAS Studio, R & Python in SAS, Analyzing Big Data
- Ogden: SAS Studio, R & Python in SAS, Analyzing Big Data
- CHHS: SAS Studio, Analyzing Big Data

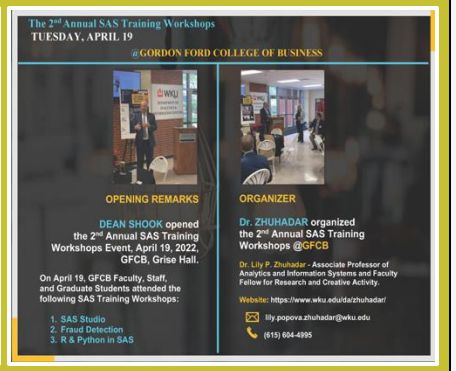
2021 Spring Break SAS & JMP Zoom Workshops

- Modeling & Machine Learning in SAS Viya
- JMP Pro for Teachers & Researchers

2019: 1st Annual SAS Training Workshops

- SAS Studio

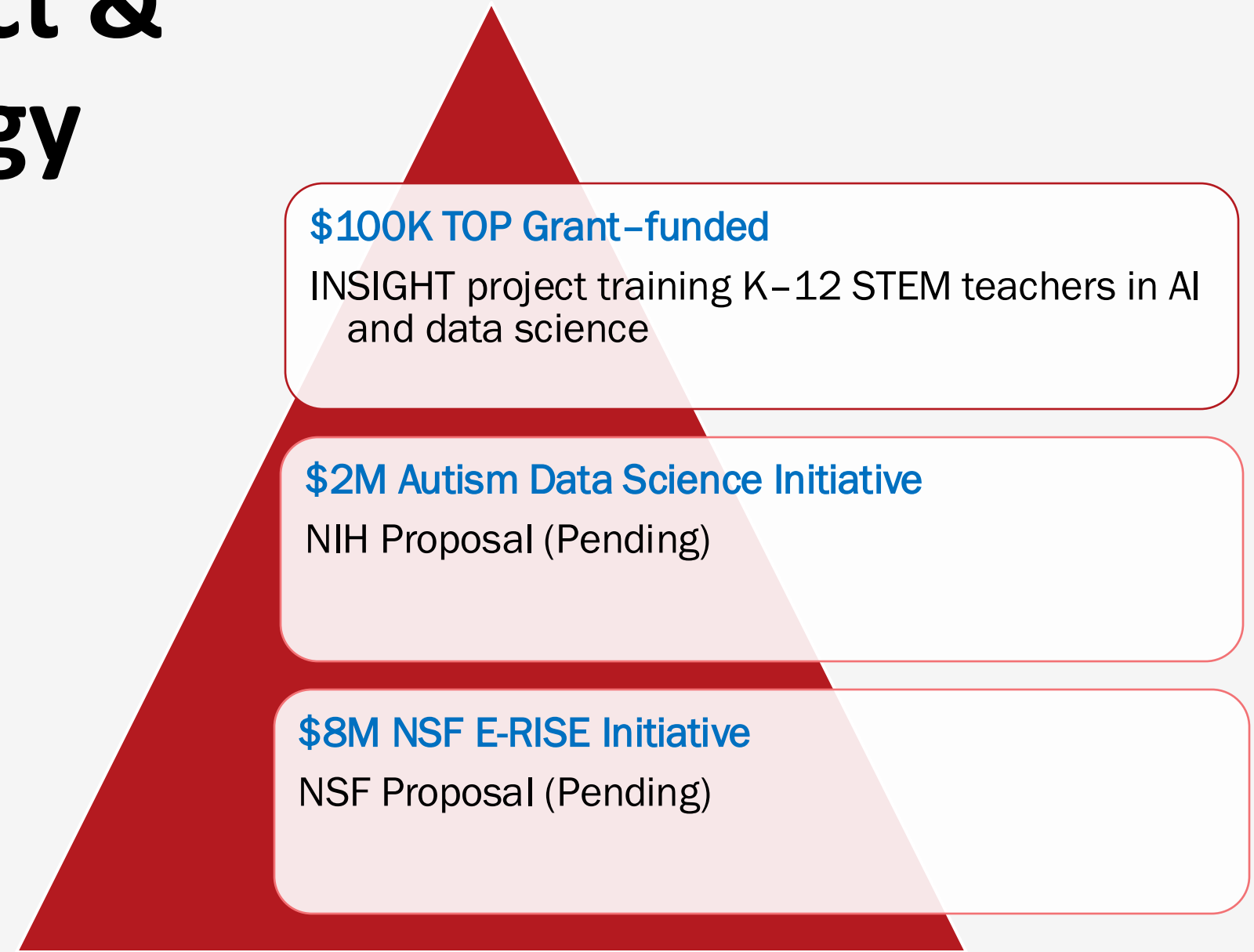




Through Business Partnerships

CADA Empowered 1,500+ Professionals and Students Through 25+ Data Analytics Training Sessions.

3.Grant Impact & Future Strategy



100K INSIGHT (TOP Grant — Funded)

Incubator for Next-Gen STEM Innovation, Growth, and Hands-on Training

- INSIGHT: Train 20 KY STEM teachers in AI/DS
- Summer institute + year-round coaching • AI/DS modules
- Team: CADA + Teacher Ed + Gatton + OIR + CITL
- Why: Grow AI-ready teachers statewide (focus on rural)
- Outcomes: Module hub • stipends & toolkits • dashboards • conference papers

Roadmap: Year 1 pilot (10 teachers) → Year 2 scale (10 more) → 600K NSF RET follow-on.

BS in Data Science @ WKU

Joint degree from Gordon Ford College of Business + Ogden College of Science & Engineering—where tech + business meet real-world impact.

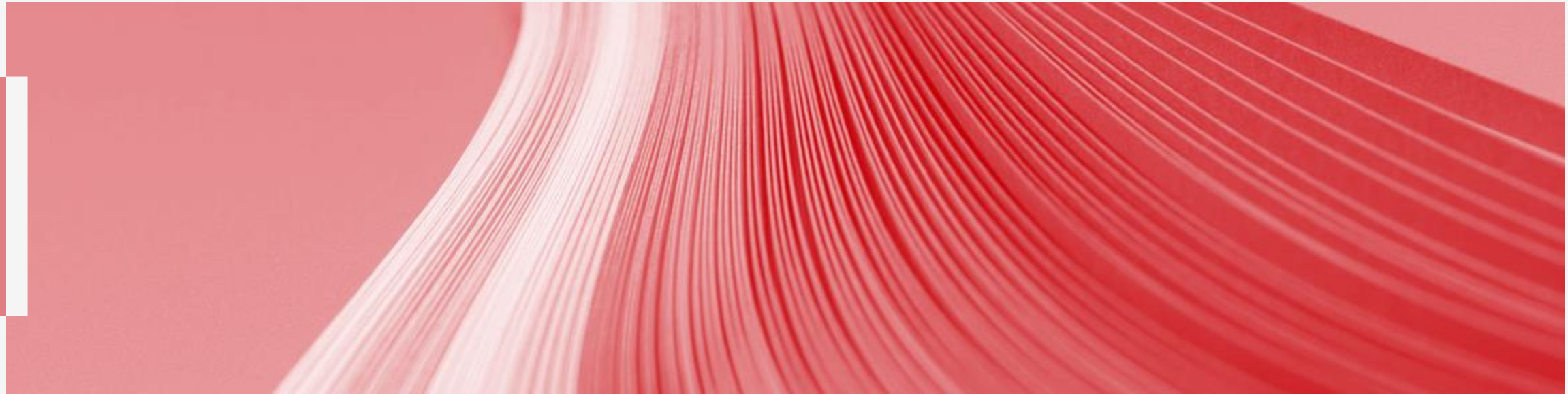
- Interdisciplinary: CS, statistics, business analytics, economics
- Design: Core DS + required minor/certificate to tailor your path
- Skills: Python/R/SQL, databases, stats/ML, visualization, business savvy
- Experience: Hands-on projects, internships, faculty-mentored capstone
- Career prep: Resume/interview coaching; strong alumni/industry network
- Roles: Data Analyst, ML & Predictive Analyst, Market Research, Data Scientist

Board Ask

Which of these can you commit to this year?

1) Sponsor AI Classroom Toolkits ($1–5K per teacher).
2) Host guest talks, site visits, summer student placements.
3) Provide placements for internships.
4) Sustainability: commit multi-year sponsorships/corporate gifts and help launch a Teacher Fellows endowed fund.

Thank you!



Cited References

Bellarmine University. (n.d.). *Challenges in 6-12 STEM education in Kentucky*. Bellarmine University. Retrieved March 11, 2025, from https://www.bellarmine.edu/education/noyce-capacity-building-project/challenges-in-6-12-stem-education-in-kentucky/

Center for Security and Emerging Technology (CSET). (2023). *The global distribution of STEM graduates: Which countries lead the way?* Georgetown University. Retrieved March 11, 2025, from https://cset.georgetown.edu/article/the-global-distribution-of-stem-graduates-which-countries-lead-the-way/

World Economic Forum (WEF). (2025). *The future of jobs report 2025*. Retrieved March 11, 2025, from https://www.weforum.org/publications/the-future-of-jobs-report-2025/

Appendix A

$2M Autism Data Science Initiative

- What: SMART trial to deliver adaptive, data-driven mental-health supports for autistic college students—adjusting intensity by early signals (GPA, attendance, CCAPS, QoL).
- Who: PIs: Christina Noel, Kim Link & Lily Popova Zhuhadar
- How: Start with campus services (SARC or KAP), then re-randomize non-responders to targeted digital tools (CBT & mindfulness apps) or human-delivered EF coaching/counseling at Weeks 5 & 10.
- Outcomes: Improved retention, mental health, and quality of life; replicable algorithm universities can adopt nationally.

CADA Role: Lead data architecture, modeling, and evaluation.

Appendix B

$8M Statewide Initiative

- What: Statewide incubator to design sustainable, AI-ready data centers—optimizing siting, power, cooling, and community impact.
- Why Kentucky: Competitive sites (incl. reclaimed mines), evolving energy mix, and Tech-Hub alignment → national testbed potential.
- Who: UofL (lead) with WKU, BCTC/KCTCS, KSTC, state agencies, utilities, and industry partners.
- Research Pillars: Energy & resiliency (hydrogen/microgrids), advanced cooling (immersion/heat reuse), LCA siting models, and a unified decision-support platform.
- Workforce: Stackable micro-credentials, internships/co-ops, teacher PD, and rapid upskilling via BCTC Workforce Solutions.

CADA Role: Leads geospatial/ML modeling.