CARTE-Enbridge Bootcamp

AI and Safety

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING
Centre for Analytics and Artificial Intelligence Engineering

Predictive vs Preventative Maintenance

- Preventative Maintenance
  - Occurs at specific intervals
  - General, blanket rules
  - Does not factor in specific conditions or changing information
  - Easy to implement
- Predictive Maintenance
  - Occurs as needed
  - Provides specific recommendations
  - Analyzes the specific conditions of equipment
  - Complex to achieve successfully

Why Predictive Maintenance?

- Improved maintenance efficiency
  - Reduces unnecessary downtime
  - Decreased labour and materials costs
- Improved worker safety
  - Better understanding of when equipment is liable to fail
  - Technicians can resolve issues before they become dangerous
- Increased productivity
  - Maintenance can be planned to factor in high demand periods, or workers' schedules

Transitioning from Preventative to Predictive Maintenance

- Initial Assessment
  - Evaluate the current state of the equipment
  - Identify KPIs
- Data Integration
  - Gather existing historical maintenance records
  - Implement sensors for real-time data collection
- Model Development
  - Use machine learning algorithms to analyze historical and real-time data
- Pilot Testing
  - Implement on a small scale, and monitor performance

Safety Monitoring

- In potentially hazardous workplace environments, following safety rules is critical
- Typically, safety officers work to ensure that workers follow safety standards properly
  - However, they can’t be everywhere, or catch every error
- AI can be used to back up human workers, notifying safety officers and workers of issues as they arise

Implementing AI Safety Monitoring

- With existing data, such as CCTV, computer vision can instantly detect dangerous situations and notify
  - Workers not wearing proper PPE
  - Undetected spills or obstacles
  - Even poor posture!
- Wearable devices can allow for spotting tiredness, exposure to hazardous substances, or even notify for potential injury
- Environmental sensors can be deployed to detect factors like safe noise levels, air quality and more

Complementing Existing Safety Procedures

- These technologies can work to assist safety officers, allowing them to better prioritize their time on the most important cases
- When a simple fix is available, employees can be notified directly
- Better safety data allows for improved understanding of area- or company-wide trends and challenges

AI Safety - Risks

- Data about individuals' behaviour and actions must be treated carefully. Challenges can be divided into:
- Non-robust
  - AI systems may work well in development, but not in practice
  - Frequent, irritating alerts about behaviour may damage safety practices instead of improving them
- Privacy violating
  - Employees' privacy must be respected, and data collected should only be used for the purpose it was intended for

AI Safety - Risks

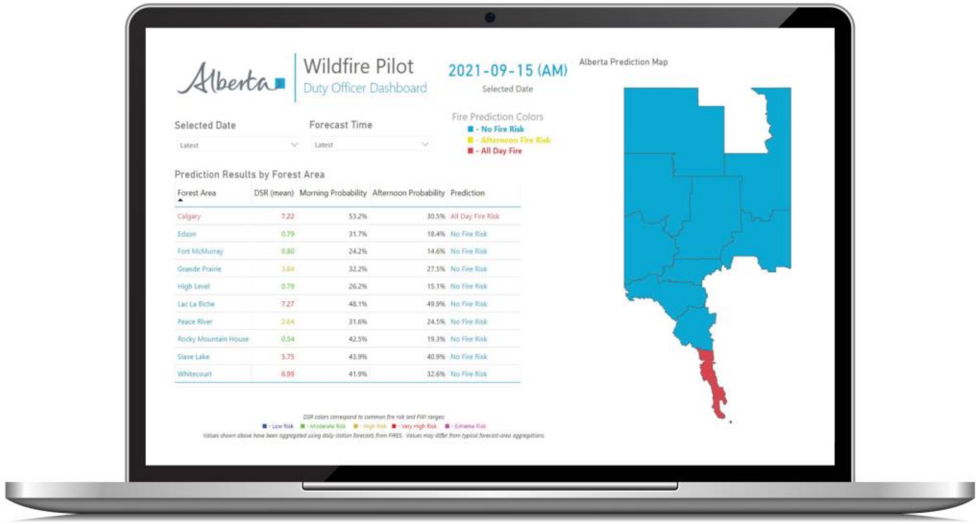
- Biased
  - When trained on biased data, AI systems can learn to replicate unwanted behaviour, such as unintentionally singling out certain groups or even individuals
- Inability to explain
  - Motivation behind a models' decision should be easy to understand, so that its recommendations are trusted

AI for Natural Disasters

- As climate change increases the occurrence of natural disasters, the value in predicting and understanding extreme weather is only growing
- Predicting extreme weather events and natural disasters allows for better preparedness, faster recovery, and safer industry

AI and Wildfire Prediction

- Last year, the AI model was deployed for the first time as part of Alberta’s toolkit to plan for wildfire season
- Cost savings are estimated at between $2m and $5m



Real-Time vs Historical Safety Prediction

Real-Time

- Immediate action
  - Algorithms trigger instant safety measures like shutting down equipment
- Sensor-based
  - Heavily reliant on IoT devices and sensors for real-time data collection
- High computing needs
  - Requires significant computational power for real-time analysis
- Narrow focus
  - Generally specialized for specific tasks, e.g. fire detection
- Low tolerance for errors
  - Mistakes can result in immediate consequences

Historical

- Strategic Planning
  - Data is used for long-term safety planning and policy development
- Database-driven
  - Utilizes archived incident reports, logs, and other records
- Low computing needs
  - Analysis can be run in less than real time, and in batches
- Broad scope
  - Useful applications can be uncovered as data is analyzed
- Iterative improvement
  - Errors can be corrected in subsequent analyses and strategies

Using Historical Data for Safety Planning

- Collecting accurate, reliable data is essential to useful modelling
- There are many types of data that can be collected
- Key concerns
  - Coverage: essential that we capture the specifics of what went wrong
  - Accuracy: particularly with human-entered data, errors can cause problems in inference
  - Privacy: we must respect the rights of employees and ensure that they feel comfortable in their workplace

Types of Safety Data

- Sensors and IoT Devices
  - Continuously monitor machinery, environmental conditions, and other factors
- Incident Reports
  - Manual entry of safety incidents, usually by safety officers or employees
- Surveillance Systems
  - Video and audio recordings can be used to analyze circumstances leading to an incident
- Operational logs
  - Automatic logs of equipment operation, maintenance schedules, and more

Types of Safety Data

- Public and Government Records
  - Information from public sources, like weather patterns, regulatory filings, can be included
- Social Media and News
  - Public sentiment and news reports can provide auxiliary information
- Interviews and Surveys
  - Sometimes, qualitative data can offer insights that are difficult to capture quantitatively
- Third-Party Data
  - Specialized companies offer datasets that can be integrated, such as geospatial data for natural disaster prediction
- Legacy Systems
  - Older data and even paper records can be digitized and analyzed

Pre-Processing Data

- Once data is collected from many sources, we need to convert it into a single useful dataset
- Begin by pre-processing
- Some common steps:
- Cleaning
  - Remove duplicates and errors
  - Handle missing values
  - Rectify inconsistencies
- Transformation
  - Normalize scales, encode categorical variables
- Integration
  - Match schema between datasets to ensure consistent representation

Data Aggregation

- Depending on the type of data, it may be necessary to aggregate to ensure utility and reduce storage requirements
  - If a sensor records every second, do we need that level of detail from 2006?
- Some options:
  - Summarization: aggregate into statistics like mean, count over time, etc
  - Time-series aggregation: convert hourly data into weekly, for example
  - Spatial aggregation: combine data from same physical sources, or locations

Explainable Models

- Many machine learning approaches are called "black-box" models
- We can verify the reliability of the predictions, but we can't explain why the model predicted what it did
- This poses a significant problem in areas where justifiability is essential

Explainable Models

- An explainable model is one whose decisions can be understood and interpreted by us
- Explainable models can still be “black-box” in that we might not be able to understand the workings directly
- But we can describe their behaviour accurately
- Example: Random Forests
  - Feature Importance
  - Partial Dependence
  - SHAP

Glass-Box Models

- Glass-Box models are <u>inherently</u> interpretable
- This makes them a type of explainable model
- They offer a high degree of interpretability, but potentially at the cost of predictive power
- Pros: directly interpretable, easier to fix
- Cons: may sacrifice model complexity for ease of understanding
- Example: Linear Regression
    - Interpretable
    - Transparent
    - Debuggable

Explainable Boosting Machines

- Combines high accuracy with easy interpretability
- Extension of decision trees
- EBM builds a distinct model for each input feature, learning its relationship to the outcome
- Then, relationships between features are automatically identified, and included in the model as new terms
- The final model combines individual feature functions with interactions
- This allows us to directly track how different features impact prediction

Explainable Boosting Machines

- Pros:
  - High Accuracy: Comparable to complex models like Random Forests.
  - Interpretable: Each feature's contribution is clear and can be visualized.
  - Fast Prediction: Quick at making real-time decisions once trained.
  - Feature Interactions: Automatically detects and incorporates how features influence each other.
- Cons:
  - Slow Training: Takes longer to train compared to some other models.
  - Computational Cost: The feature-by-feature training and interaction detection can be computationally intensive.

What’s Next?

- Improved real-time monitoring
  - More and better sensors allow for improved predictions
- Worker health monitoring
  - Wearables allow for immediate identification of health risks
- Explainable AI
  - As an active area of research, improvements in modelling allow for ‘best of both worlds’ approaches

Summary

- AI in Maintenance and Safety
  - Transition from Preventative to Predictive Maintenance for efficiency and safety.
  - AI-backed safety monitoring complements human efforts in real-time.
- Explainable AI
  - High-accuracy models that maintain transparency and interpretability.
- Future Trends
  - Real-time monitoring, worker health insights, and advancements in explainable AI.
  - Expanding into various sectors like agriculture, healthcare, and manufacturing.

Takeaways

- Efficiency and Safety
  - AI can significantly improve both efficiency and safety in industrial settings.
- Human-AI Collaboration
  - AI systems can work alongside human workers, not replace them.
- Transparency is Key
  - As AI models get more complex, making them understandable becomes crucial.