Data Insight: Concepts and Case Studies

for a General Audience

EUROPOL

Overview

Day 1: Data Insight

- Policing with Data & Case Studies in UK Policing
- Data Insight
- Programming & Automation
- Roundtable Close

Day 2: Big Data & AI

- Big Data
- AI
- Machine Learning
- Natural Language Processing & Ethics

Day 3: Mindset & Practice

- Data Mindset & Storytelling
- Data Ethics
- Data Governance
- Digital Transformation: Roundtables & Close

Day 1

Intro & Roundtable Questionaire: What interests you?

- Introductions
- Starting Questionaire

Policing with Data & Case Studies in UK Policing

- Case Studies in Problems
    - Challenges in Data Collection
    - Case Study: House Burglery & Data Siloing
- WB. Recording a Crime
    - Review: Recording the Crime
    - Review: Intelligence Gathering
    - Review: Arrest
- Challenges in Data Quality
    - Problem: Structure
    - Problem: Field Structure
- Problems
    - Problem: Uniqueness
- Challenges in Modelling
    - Problem: Behavioural Change
    - WB. Future = Past?
    - WB. Criminal Network Analysis
        - Review: Problem: Network Analysis
    - Problem: Priotisation of Offenders
        - WB. Factor Analysis

- Challenges in Research & Experimentation
  - Problem: Getting Data
    - Problem: Unifying and Annoymising Data
    - Projects: Europol EPRIS
  - Problem: Decision Quality
  - Break Exercise
- Case Studies in Solutions
  - Case Studies: Supervised Learning
    - for Interview Crime Lists
  - WB. Supervised Learning
  - Case Studies: Unsupervised Learning
    - for Early Detection of Crime Series
  - WB. Unsupervised Learning

Data Insight

- Discuss: Why now?
    - Case Study: Predictive Policing
    - Where does the insight boom come from?
    - When did data insight start?
    - When did organizations start being data-driven?
- What is Data Analysis, Science and Insight?
    - What are traditional data practices?
    - What are the new insight practices?
- What is Data?
    - What is Insight?
    - What is inference?
    - How do you get Data Insight from Raw Data?
- What is a typical data insight workflow?
    - What techniques and technologies support Insight?
    - What tools does data insight require?
- What roles do data insight functions require?
    - How do you set up a data insight project?
    - What questions should I ask of a data project?
    - How do data practicioners solve problems?
- WB. What is Probability?
    - Review: What is probability?
    - WB. Case Study: Forensic Risk Analysis
- Group Discussion & Reflection

Programming & Automation

- What is Automation?
    - How do organizations get an ROI on automation?
    - What types of Insight can Automated?
    - Why replace human-driven applications with programs?
    - WB. What tools do analysts & scientists use?
    - Review: What tools do analysts and scientists use?
- Review: What is Python?
    - Demo: How do you write a program?
    - Demo: How do you use libraries in a program?
    - Review: What are the advantages of python?
    - Review: What is Programming?
    - Review: How do you program Insight?
- Reflection
    - Individual Project

Day 2

Introduction to Big Data

- Why now?
    - Why is Big Data important?
    - WB. Who cares about Big Data?
- What is Big Data?
    - How do we decide if a problem is Big Data?
    - WB. When is a problem Big Data?
    - WB. Are all Big Data Systems for Big Data?
    - WB. What is a relational database?
- The 3 Vs
    - Velocity
- Vairety
    - WB. Variety: Relational & Non-Relational
    - WB. Variety: K-V, Documents
    - WB. Variety: Graphs
    - WB. Variety: Columnar
    - WB. Variety: Images & Audio
- Volume
- Review: What is a Data Model?
- Review: What types of Big Data Systems are there?
- WB. Data Archicture & Components
    - What does a modern data system need?
    - Case Study: NoSQL vs. Relational
- Individual Project (10 min)
- Group Project (20 min)

- Appendix
    - What is SQL?
    - How do big data systems work together?
        - What is a Traditional Data System?
        - When do traditional systems win?
        - When do traditional systems fail?
        - When is a big data system needed?
    - What roles enable big data?

Introduction to Artifical Intelligence

- What is AI?
    - WB. What is Automation, Augmentation, Assistance...
    - WB. What is AI? Weak vs. Strong
    - WB. What is Weak AI?
    - WB. What is Strong AI?
    - WB. What Weak Systems exist?
    - WB. How have weak systems developed?
    - Review: Weak AI, Strong AI
    - Review: When did "AI" start?
    - Review: What are symbolic AI systems?
    - Review: What are statistical AI systems?
    - Review: What is the value of AI?
- Individual Project: (10 min)
- Group Project (20 min)
    - WB. Review: AI Adoption
    - WB. Review: AI Barriers

- Reflection
  - What are the challenges to AI adoption?
  - What is Weak vs. Strong AI?
  - What is Narrow Intelligence?
  - What is General Intelligence?
  - Is General Intelligence a form of Narrow Intelligence?
  - Saftey
    - How do we improve the safety of ML?
    - What does intelligence?
  - Robotics

Applying Machine Learning

- Review: What is AI?
- What is Machine Learning?
    - What can Machine Learning do?
    - WB. The types of Machine Learning
    - WB. How does Machine Learning work?
    - WB. Regression
    - WB. Classification
    - WB. Clustering
    - Demo: How do you create machine learning systems?
    - Review: How does Machine Learning work?
- When can Machine Learning add value?
    - Demo: sklearn Classifier Comparison
- What is Deep Learning?
    - Where is the business value in Deep Learning?
    - Deep Learning: What is Natural Language Processing?
    - Deep Learning: What is Computer Vision?
    - Demo: AI as a Service: How can you deploy Deep Learning?
- Case Studies
    - Case Studies: Supervised Learning
        - Automated Facial Recognition
        - Voice Identification
    - Case Study: Forensic Unsupervised Learning
    - Predictive Identification
        - Individual & Group Project

Natural Language Processing & Ethics

- How are words represented to machines?
    - WB. How do you compare words?
- How do machines process Dostoevsky?
    - Let's get Crime and Punishment
    - Let's model a vocabulary
    - Let's find out how Dostoevsky uses words
    - Does the distributional hypothesis work?
- Natural Languages
    - What do words mean?
    - Can we learn word meanings by parsing text?
    - What can we learn from text?
    - What is association?
    - Are Human Implicit Associations expressed in Text?
    - Is association part of semantics?
- Bias
    - What is Bias?
    - WB. How do associations relate to prejudice?
    - Review: How do associations relate to prejudice?
    - Can machines detect stereotypes?
    - Where does AI bias come from?
    - Is AI racist just because we are?

Day 3

Data Mindset & Skills

- Why do you need a Data Mindset?
    - Discuss: What are the components of a Data Mindset?
    - Review: What are the components of a Data Mindset?
    - What mindset does a data practicioner need?
    - What mindsets does an organization need?
- Reasoning
    - What is Reasoning?
    - What is(n't) Evidence-Based reasoning?
        - Superstition
        - Speculation
        - Evidence-based reasoning
    - What techniques and technologies support evidence-based reasoning?
    - What are cognitive biases?
        - Anchoring and Adjustment Effect
        - Clustering illusion
        - Framing Effect
        - Availability heuristic
        - Base rate fallacy
        - Selective perception
        - Group think
    - How do you formulate a hypothesis?
    - How does data distinguish between hypotheses?
    - Demo: Amnes Room

- Storytelling & Visualization
    - What is Data Storytelling?
        - Project Valcri
    - What is a story?
    - Why are data projects presented as stories?
    - What is Visual Science?
    - How do you create good visuals?
    - DEMO: Ted Talk
- Group Project: Write a Data Story (30 min)

Data Ethics

- Why Data Ethics?
    - What is Data Ethics?
- Policing Case Studies: Motivations
    - Data Storage
    - Illegal Actions
    - Bias
- Mini Case Studies: Motivations & Problems
    - Case Study: Tesla Crash (Responsibility)
    - Case Study: `care.data` (Importance of Ethics)
    - Case Study: Airport Threat Detection (Risks)
    - Case Study: Unintended behaviour
    - Case Study: Street Bump (Inequality)
- Topics in Data Ethics
    - Ethical Systems
    - The Ethics of (Opaque) Algorithms
    - Problem in Data Ethics
        - Problems: Privacy
        - Problems: Discrimination
        - Problems: Auditing & Responsibility
- Topic Focus: Accountability & Explanation
    - Case Study: Compass Recidivisism System
    - Why don't we get explanations?
    - What is an Explanation of a Model?
    - What is an Explanation?
    - Counterfactual explanations of models
- Group Exercise

- Appendix
  - The Law
  - Concerns
    - Changing Laws
    - International Regulartory Concerns
  - GDPR
  - Data Protection Directive (Not Legally Binding, presently) -- Art 12a
  - Art 22 GDPR -- Automated Individual Decision-Making

Data Governance

- What is Data?
- What is Data Governance?
    - WB. How do you manage data as an asset?
    - WB. What does Data Governance oversee?
    - What are the goals of Data Governance?
    - WB. What areas does governance touch?
    - Review: What topics does governance cover?
- How does Data Governance relate to Compliance?
    - Governing Organizations
- How do we govern data systems?
    - How does meta-data help govern data systems?
    - WB. How are Data Governance roles defined?
    - How do organizations implement Governance?
    - What is Data Stewardship?
- The Data in Data Governance
    - How do you value data?
    - How do we ensure data quality?
    - How do you manage data security?
- Group Exercise

Digital Transformation & Data Insight

- Overview
    - WB. What have we covered so far?
    - Reflection: The topics so far
- Data Insight Rewards & Risks
    - What are the barriers to adopting insight practices?
    - What are the strategic advantages from data insight?
    - What are the risks of data insight projects?
    - When will data insight be rewarding?
- Group: Data & Digital Transformation
    - WB. Risks & Rewards
    - What are the organizational components of an insight culture?
    - WB. The Organizational Structure for Digital Transformation
- Data Insight Organizational Questions
    - What is a data democracy?
    - What do workers want from a data democracy?
- Reflection: How do you implement a digital transformation?
    - Roundtable
        - WB. Roundtable
- Case Study: When will data insight help?

Live Demonstrations

- Fundamental Concepts in Data Insight: **Demo: NoSQL & Non-Relational Data**
  - What is a Key-Value model?
  - What is a Document model?
  - What is a Graph Data Model?
  - What is a Columar Model?
  - Appendix
    - Hashes and Encodings

- Fundamental Concepts in Data Insight: **Demo: Relational Data & SQL**
  - How do you define data schema?
  - How do you insert data?
  - How do you query for data?
  - How do you peform analytical queries?

- Fundamental Concepts in Data Insight: **Demo: Automating Insight**
  - What are algorithms?
    - How do you write an algorithm in python?
  - How do you write calculative algorithms?
  - How you access pre-defined algorithms?
  - How do you combine business-specific and generic algorithms?

- Fundamental Concepts in Data Insight: **Demo: Automating Insight**
  - The Simulation
  - Descriptive Analytics
  - How do you automate insight?
  - Exercise

- Fundamental Concepts in Data Insight: **Demo: Machine Learning Concepts**
  - What dataset are we using?
  - How do I query this dataset?
  - How do I predict probability of a stop-and-search success?
  - What's a k-NN query?
  - How would we automate this analysis with python?
  - Summary & Review

- Fundamental Concepts in Data Insight: **Demo: Unsupervised Learning & Network Science**
    - Data
    - Aside: Visualizing the Hierachy

- Fundamental Concepts in Data Insight: **Demo: AI & Deep Learning**
  - What data set are we analysing?
  - What is sentiment analysis?
  - What messages does the dataset contain?
  - What sentiments does the dataset contain?

Appendix Resources

Case Studies in Police Data Systems

- General Observations
  - General Trends
  - General Innovation Topics
  - General Policing Areas
  - General Ethical Topics
  - General Organizational Challenges
  - General Cultural Challenges
  - General Regulatory Challenges
  - General Technical Challenges

- Case Studies & Applications
  - Data Collection
  - Boders & Querying
  - Automated Licence Plate Readers
  - Automated Facial Recognition
  - Voice Identification
  - Predictive Mapping
  - Predictive Identification
  - Databases
  - Hetrogenous Databases
  - Mixing Data Systems
  - UK DB Interoperability
  - DB Proliferation
  - Europol Databases
  - Ethics
  - Data Storage
  - Law breaks the Law
  - Extra-Legal Actions (no legal frameworks)
  - Social Groups & Privacy
  - Bias

- AI & Insight
    - Predictive Policing
    - Predictive Mapping
    - Predictive Identification
    - Real-time Identification
    - Automated License Plate Readers
    - Automated Facial Recognition
    - Voice Identification
    - Regulatory Challenges
    - Changing Laws
    - International Regulartory Concerns

- Articles
    - References
    - Further Reading: General Topics
    - Further Reading: Projects
    - Further Reading: Law & Ethics
- News
- Videos
- Tools