Data Analysis

- Inspect, clean, transform, model data to discover useful insights.

- Manual methods & statistical techniques for patterns.

AI enhanced DA

- leverage ML, NN, automated algos to analyze vast amts of data.

- AI + DA = faster, accurate, insightful, scalable

AI

- algos + ML to simulate human intelligence in computers for
  - complex problems
  - pattern recognition
  - data driven decisions

Cognitive comp
CV
ML
Deep learning
NLP

ML : learn from data w/o explicit pgm. improve over time

DL : subset of ML ANN trained on large datasets via supervised / unsupervised.

NLP: interpret & generate human lang.

CV: use of algos to interpret visual data.

CC: simulate human thought process w/ ML.

Why?

① Exponential growth of data
  - IoT, social media
  - trad methods limited (effort, time, comp^n)

② Traditional analysis kinda sucks.
  - requires humans (time, error ↑)
  - large, unstruc, complex data = hard
  - scalability + realtime = issue
  - requires skilled analyst.

What can AI do differently?

→ Scalability (run job at given time — automatically)
→ Automation (reduce human error, improve efficiency)
→ Enhanced insight (hidden patterns that humans miss)
→ Query processing (just make an agent)

Benefits of AI in DA

① Speed + efficiency :
  a) II^d processing + optimized algos
    - multiple ops simult.
    - super high speed

  b) Real time analytics
    - for massive datasets
    - prompt responses to changing demands.

  c) Data volume scalability
    - can handle large amts of data w/o line by line review
    - reduces bottlenecks.

② Continuous improvement
  a) Pattern recognition w/ ML
    - NN + gb algos
    - identify intricate, non-linear patterns

  b) Error minimiz^n thro' model refinement.
    - refined w/ backprop, cross val
    - ensure high accuracy

  c) Data cleaning
    - identify inconsistencies, outliers, anomalies
    - reliability of processed data ↑

③ Automation of repetitive tasks.
  a) Automated data prep

Challenges of using AI

① Data privacy + security
  a) Automated data prep
    - privacy risk w/ large vols of data
    - GDPR + CCPA

  b) Data storage
    - protect data during AI model training
    - data encrypt, secure multi-party comp^n, differential privacy

  c) Data anonymization + pseudonymization
    - to mask personal identifiers

② Bias
  a) Data bias + representation
    - biased data = biased model.

  b) Algorithmic + model bias
    - unintended bias in algos = :(

  c) Mitigation strategies
    - resampling, adversarial debiasing, fairness aware learning

Namita Achyuthan

③ Automation of repetitive tasks.
   a) Automated data prep
      • ETL (extract, transform, load) pipelines
      • min. human intervention

   b) Text + document analysis
      • integrate LLM + doc. parsing pipelines

c) Mitigation strategies
   • resampling, adversarial debiasing, fairness aware learning

③ Need for human oversight
   a) Complexity in interpretability
      • black box — hard to interpret.
      • validate AI-driven conclusions

   b) Human expertise for contextual understanding
      • AI = no knowledge for contextual decisions

   c) Ethical + regulatory compliance
      • ensure AI models adhere to standards.

LLMs

- Large Lang. Model : neural net w/ multiple parameters.
  ↓
- Trained on large vols of text : SSL.
  ↓
- LLM understands patterns, grammar, context

    └→ LLM + DA
        ↓
      adv. lang processing + insight from large vol. of data

DA in business

① Descriptive — historical data for patterns

② Diagnostic — cause of past outcome

③ Predictive — forecast future trends

④ Prescriptive — recommends actions

① NLP for unstructured data
   - significant portion of business data is unstruct.
   ✓ enhanced text analysis
   ✓ automated summarization

② Improved data interpretation
   - LLMs can provide clear explanations
   ✓ natural lang. expl^n
   ✓ contextual insights

③ Automated reporting + dashboarding
   - manual generation = time consuming
   ✓ dynamic report generation
   ✓ interactive dashboards

④ Predictive analysis + forecasting   (RETAIL)
   - process historical data + forecast based off historical trends + patterns
   ✓ data driven forecasting }(FINANCIAL
   ✓ scenario analysis   INSTITUTIONS)

⑤ Enhanced customer insight
   - understanding customer behavior is v. impt.
   ✓ Sentiment analysis (HOSPITALITY)
   ✓ Personalized recommendations

LLM for DA Workflow

LLM for DA Workflow

Project Scope + Goal

① Identify Use Cases
   - determine specific DA tasks →
     - generate insights
     - automate data querying + retrieval
     - analysis

② Set clear goals
   - define what success looks like
   - measurable outcomes

Data collect⁰ and prep

③ Data Collection
   - Types of data for training →
     - struct (db, xlsx)
     - unstruct (report, mail, posts)

④ Data quality
   - high quality + relevant + diverse

⑤ Data preprocessing
   - handle missing vals
   - normalize
   - tokenize

Finalize model arch.

⑥ Choose base model

⑦ Customize architecture
   - depends on use case

Train, evaluate, deploy.

⑧ Train
   - suitable env (hardware + framework)
   - finetune n/ specific datasets

⑨ Evaluate
   - test on test dataset
   - evaluate generalization

⑩ Deploy
   - cloud?
   - existing workflow?