Lessons from the Future

Predictions in Finance, Sports and Medicine

Dr. Robert P. Schumaker
Associate Professor of Computer Science
Director – Data Analytics Lab

The University of Texas at TYLER

Lessons from the Future

- What is a Data Scientist?
  - Solve large, complex, often unstructured or ill-defined problems using a data-driven approach
  - Able to predicting customer needs and anticipate orders
  - Someone that sees patterns, subtle details, looks at the world differently

Lessons from the Future

- Is Data Science just Machine Learning?
  - Machine Learning is a part of the process
  - 1. Must understand the problem
  - 2. Must know what ML algorithm to use based on its strengths and limitations
  - 3. MUST BE ABLE TO INTERPRET THE RESULTS

Lessons from the Future

- What is Data Science?
  - Knowledge of Statistics – Statistician
  - Clean and Analyze Data – Business Analyst
  - Able to Data Mine – Entry-level Data Scientist
  - Translate Business Requirements – Data Scientist
  - System Builder – Senior Data Scientist

Lessons from the Future

- The Data Science Approach
  - Identify a general area to study
  - Acquisition
    - Determine what data resources are available
    - Collect data directly or through API
    - Clean the relevant data
  - Representation
    - Ascertain the independent and dependent variables
    - Shape the research questions
    - Consider ways to model the data

Lessons from the Future

- The Data Science Approach
  - Description / Prediction
    - Determine the relevant statistics, machine learning algorithms
    - Train, Test and Validate
    - Refine research questions, seek additional data sources, iterate
  - Delivery
    - Prepare a technical research report to communicate results
    - Present the findings
    - Create a software prototype

Lessons from the Future

- Lessons in Finance (Textual/Financial Prediction)
  - How it started

THE WALL STREET JOURNAL.

VOL. CCX NO. 79 — TUESDAY, OCTOBER 20, 1987 — 50 CENTS

Risky Strategy

Attack on Iranian Site Wins Praise but Means Violence May Escalate

U.S. Says Strike Shows It Is In Persian Gulf to Stay; Congress Remains Uneasy

Soviet Role Is Called Unlikely

What's News—

Business and Finance

World-Wide

Market Fallout

The Market Debacle Rouses Worst Fears Of Little Investors

Many Are Beginning to Talk In Terms of Doomsday; Big Investors Are Braver

The Crash of '87

Stocks Plunge 508 Amid Panicky Selling

Percentage Decline Is Far Steeper Than '29; Bond Prices Surge

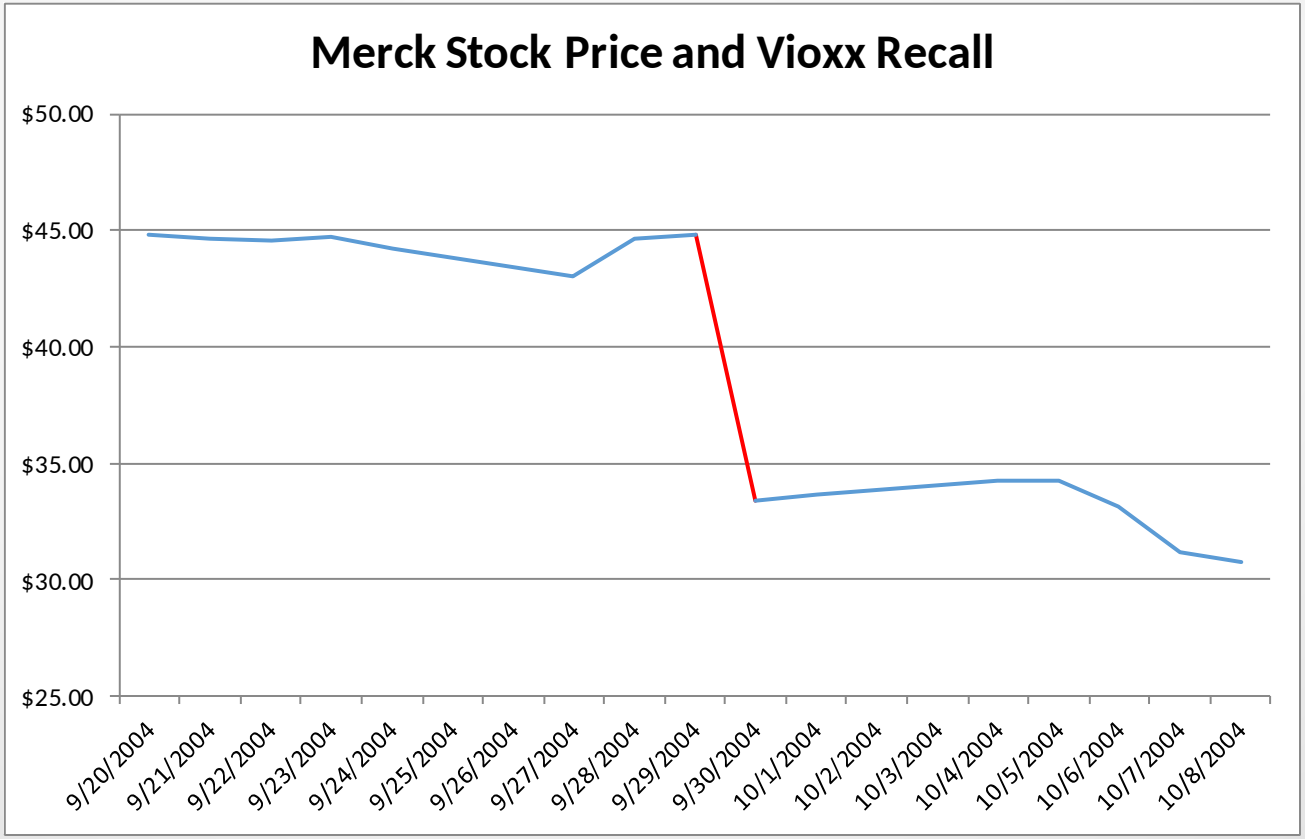
The Big Crash

A Repeat of '29? Depression in '87 Is Not Expected

Banking System Safeguards And Federal Mechanisms Are Viewed as Adequate

Lessons from the Future

- Textual/Financial Prediction
  - Building a news-aware trader

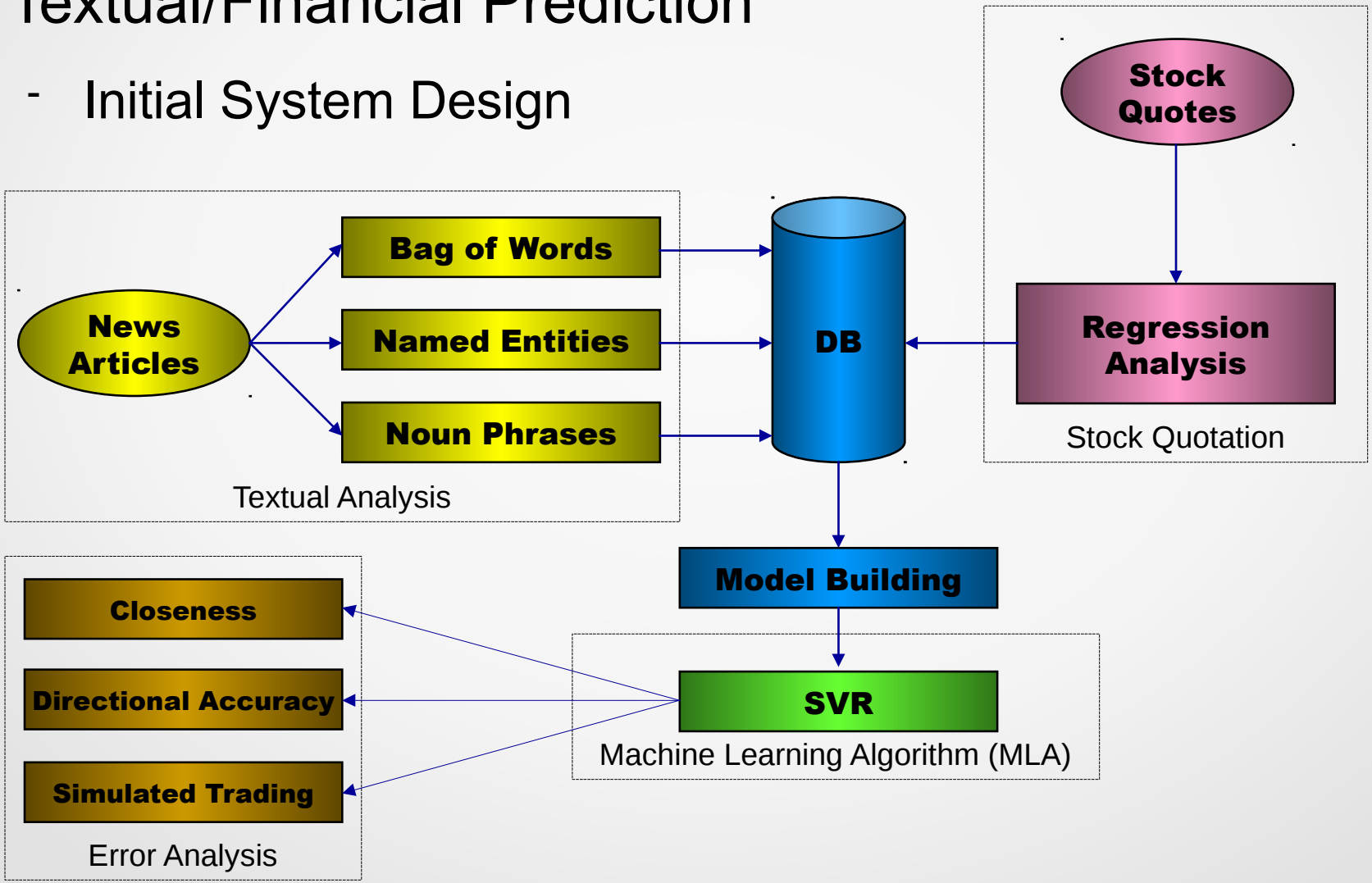


Lessons from the Future

- Textual/Financial Prediction
  - Problems
    - Prior research focused on price direction prediction / classification
    - Prior research used only Bag of Words
    - Very limited work, lacked critical details
    - No research built a simulated trading engine (HFT)

Lessons from the Future

- Textual/Financial Prediction
  - Initial System Design



Lessons from the Future

- Textual/Financial Prediction
  - Later System Designs
    - Similarity in industry classification
    - Comparison to trading experts
    - Building a portfolio
    - Adding dimensions of sentiment
    - Stock Price Management
      - Should the article be released
      - When should it be released
      - What media should it be released in
      - *What terms should be used*

Lessons from the Future

- Textual/Financial Prediction
  - Lessons learned
    - Data Science takes time
    - Need to be cross-disciplinary
    - Data-driven leads to unexpected insights
    - Refining variables and understanding a entire solution landscape can take a lifetime

Lessons from the Future

- Lessons in Sports (Sport Prediction)
  - How it started

Lessons from the Future

- Sport Prediction
  - Problems
    - Prior research focused on binary classification
    - Prior research manually curated inputs
    - Very limited work, **but had the critical details**
      - Small tweaks to variables, algorithms and domains

Lessons from the Future

- Lessons in Sports (Wagering Prediction)
  - How it started

#nufc

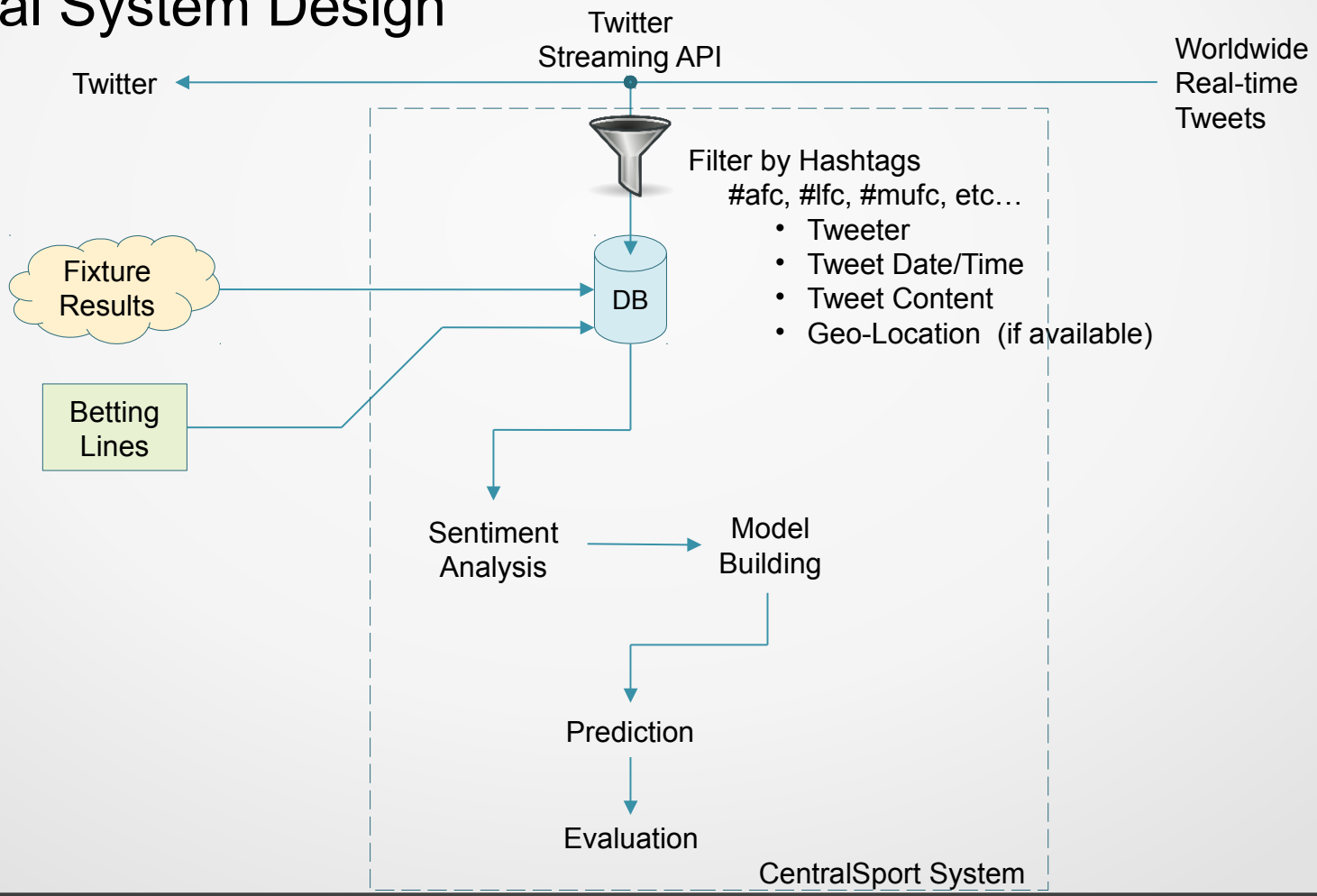
Supporting Newcastle is actually making me hate football.

Lessons from the Future

- Wagering Prediction
  - Problems
    - Prior research focused on binary classification
    - Prior research manually curated inputs
    - Very limited work, lacked critical details
    - No correlation to wagering

Lessons from the Future

- Wagering Prediction
  - Initial System Design



Lessons from the Future

- Sports and Wagering Prediction
  - Lessons learned
    - Need to be cross-disciplinary
    - Data-driven leads to unexpected insights
    - Refining variables and understanding a entire solution landscape can take a lifetime
    - Tweeters from Sunderland are not nice people

Lessons from the Future

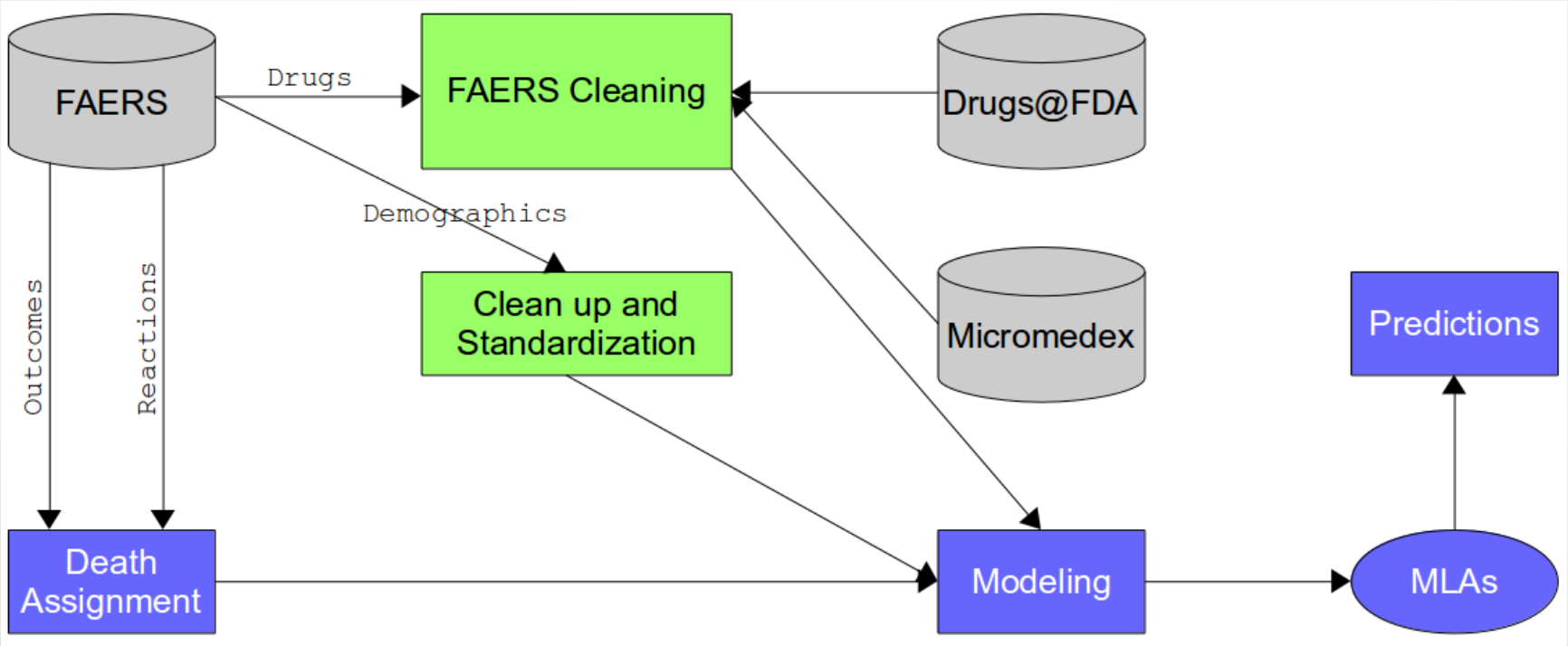
- Lessons in Medicine (ADE Prediction)
  - How it started

Lessons from the Future

- ADE Prediction
  - Problems
    - Prior research focused on 2 drug combinations
    - Available data is very large and dirty
    - No related work, lacked critical details

Lessons from the Future

- ADE Prediction
  - Initial System Design



Lessons from the Future

- ADE Prediction
  - Current System Designs
    - Look at reactions in combination with opioids
    - Partition based on patient demographics
    - Partition based on disease

Lessons from the Future

- ADE Prediction
  - Lessons learned
    - Data Science takes time
    - Need to be cross-disciplinary
    - Data-driven leads to unexpected insights
    - Refining variables and understanding a entire solution landscape can take a lifetime
    - Be flexible in system building
    - Don't take Heroin and Buprenorphine together

Lessons from the Future

- Lessons I've Learned on Becoming a Data Scientist
  - Learn as much as you can
  - Keep up-to-date
  - Don't be afraid of hard problems
  - There are no shortcuts, it will take time
  - Go where the data takes you
  - Make it fun