Overview:

The purpose of this course is to provide students a greater appreciation, both conceptual and technical of the uses of big data and its engineering of financial analytic. The course covers a number of themes and practical problems that combine computer aided data management, data analysis, financial issues and their applications in a data based financial environment.

Schedule of Classes:

<table>
<thead>
<tr>
<th>Week</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Market Data &amp; Bloomberg API</td>
</tr>
<tr>
<td>2</td>
<td>Relational Database &amp; Pair Trading</td>
</tr>
<tr>
<td>3</td>
<td>Big Data and Hadoop</td>
</tr>
<tr>
<td>4</td>
<td>Microsoft Azure Cloud &amp; Hortonworks Hadoop</td>
</tr>
<tr>
<td>5</td>
<td>Data Mining &amp; Text Miners</td>
</tr>
<tr>
<td>6</td>
<td>Neural Networks and Genetic Algorithms</td>
</tr>
<tr>
<td>7</td>
<td>Python for Finance</td>
</tr>
<tr>
<td>8</td>
<td>Final Exam</td>
</tr>
</tbody>
</table>

Assessment:

Students will have homework assignments and final exams. The assessment will be done as the following:

- Quizzes 20%
- Homework Assignment, 20%
- Lab, 10%
- Final Exam, 50%

Textbooks:


Course Topics:

1. **Financial Data:** Due to technical advancement and industry innovations such as high frequency trading, the volume of financial data will continue growing exponentially. Topics covered in this unit:
   - What are Financial Data?
     - Equities: stock quotes, daily open, high, low, close prices, trading volume, VWAP and IPOs.
     - Companies: key company fundamentals, financial statements and ratios, analyst stock recommendations and ratings, analyst earnings estimates, and financial, economic and business news.
     - Mutual Funds & ETFs: historical and daily end-of-day closing NAVs, ETF, open ended mutual funds and money market fund data.
     - Fixed Income & Credit: bond price and yield data, US Treasury and LIBOR-based swap and forward rates, and global interbank interest rates and the official BBA LIBOR.
     - Futures & Options: daily end-of-day prices, quotes, volume and open interest for all options and futures.
     - Forex & Metals: quotes for currency exchange rates, precious metal spot prices, and global currencies pairs.
     - Indices & Markets: index values for more than 10,000 U.S. and international indices, and global economic calendars of scheduled Treasury and economic events and announcements.
   - How are Financial Data Delivered?
     - Delivery frequency: real-time, delayed, conflated, or end of day (EOD).
     - Delivery method: streaming format, snapshot files or EOD files.
• Delivery transportation: broadcast, multicast, satellite, private line, VPN, or Internet.
• Delivery format: encoded heavily to optimize performance, or in simple formats to simplify databasing.
• Normalization: a vendor collects from sources all around the world and then translates all of those formats into a single format.
• Reliability: high availability of data is a primary concern in the financial markets.
• Value Added Services: data value can be improved by adding on related services such as listing information, share data, fundamental data, etc.

o How Could We Receive Financial Data?
Marketing Data Application Programming Interfaces (API): the market data from data vendors provide data-neutral, thread-aware access for requesting and providing market information as well as some generic configuration and logging capabilities. For example, Bloomberg Open Market Data Initiative, and Reuters Market Data System (RMDS) and Open Message Model (OMM) which offers an ultra low latency and high throughput data delivery option.

II. **Relational Database & Pair Trading:** This unit will introduce students to the concepts of relational database management systems and relational database model. Students will learn how to use relational database in trading via the pair trading implementation.

Topics covered in this unit:
  o Relational Database
    • DBMS
    • Relational Database Model
    • Relational Algebra
    • Normalization
    • E-R Relationship
    • Structured Query Language
  o Pair Trading Implementation
    • Microsoft JET Engine
    • Data Access Technologies
    • Pair Trading System
      • Condition and Assumption
      • Pair Trading Algorithm
      • Database Implementation Details

III. **Big Data in Finance:** While the industry structured data is growing in size and scope, it is the world of unstructured data that is emerging as an even larger and more important data source. IBM Big Data Work survey indicates that most financial organizations are currently in the early stages of big data planning and development efforts.
Topics covered in this unit:
  o What is Big Data?
  o Big Data Challenges in Financial Markets
  o Structured vs. Unstructured Data
  o Main Big Data Technologies & Hadoop
  o Harvest Financial Information Using Big Data
    ▪ Hortonworks Sandbox and HDP
    ▪ Aggregating 10 Years of Raw Stock Ticker Data from NYSE
    ▪ Enriching the Data Model with Unstructured Data from Internet
    ▪ Interactive Visualization

IV. **Data Mining in Finance**: This section will bring the concepts of data mining and data science concepts to students. Data mining algorithms and tasks will be introduced to students with focusing on text mining.

Topics covered in this unit:
  o Why Do We Need Data Mining? What is Data Mining?
  o Data Mining Algorithm
  o Data Mining Models and Tasks
  o Data Mining Development
  o What is Text Mining?
  o Text Mining Strategies
  o STATISTICA Text Miner
  o Classifying Reuters Data Set for “Earning” and Make a Predictive Model

V. **Data Mining with Neural Networks and Genetic Algorithms**: This section will describe the specifics and difficulty of data mining in financial market. Students will learn the problem definition and the frameworks for forecast performers. Students will be introduced the data-based decision-making methods for financial tasks without a model, but with sufficient training data to discover hidden regularities, with focusing on neural networks and genetic algorithms.

Topics covered in this unit:
  o Financial Problem definition
  o Learning Paradigms for Data Mining
  o Introduction to Neural Networks
  o Neural Networks and Hybrid Systems in Finance
  o Recurrent Neural Networks in Finance
  o Modular Networks and Genetic Algorithms
  o Genetic Algorithms and its Implementation