

CSCI568

Discussion 2: What is Data Mining?

Bank of America, Agoura Hills, CA

Consultant-Data Mining/Whse Finance

APPLY

Detailed Description

Participates in design, development and implementation of complex data management, storage applications, often using new technologies. Serves as a fully seasoned, proficient technical resource. Participates as an individual contributor on projects, completing activities as a part of a team related to special initiatives or operations. Routine accountability is for technical knowledge and capabilities as a team member or as an individual contributor. Works under minimal supervision, with general guidance from more seasoned consultants or managers. Work leadership may be provided by assigning work and resolving problems. Typically 5-7 years of IT experience. 6+ years of overall experience in the IT industry with emphasis on data warehouse/reporting/data analysis 4+ years of SQL Based Reporting and Analytics with strong emphasis on ETL tools, analysis and reporting services High level of SQL / RDBMS skill set with expertise in queries/stored procs/triggers/schema design - transactional and reporting High level of numerical analysis skills and ability to discern patterns in data. Excellent interpersonal communication skills Ability to run projects independently and work with a team of architects and database developers.

Compensation

Unspecified

Bank of America 

Active Jobs on JIBE: 3704

**7,092,415 JIBE users
connected to people at Bank
of America**

[See all jobs listed by Bank of America](#)

No!

Sr. Software Development Engineer- Data Mining/ Machine Learning / Information Retrieval

Job Description

Software Dev Engineer, Information Retrieval / ML / **Data Mining**

Amazon.com's Product Ads team is looking for exceptional software engineers to develop algorithms and build systems to solve a variety of information retrieval, machine learning and **data mining** problems related to the Amazon Product Catalog using large scale distributed systems.

What we do:

Our team develops and employs innovative, cutting edge techniques in **data mining**, information retrieval and distributed computing to mine the Amazon product catalog to identify relationships between various offers provided by advertisers. We develop systems that build the intelligence of the Amazon's selection, which is the Earth's largest while maintaining a relentless focus on the shopping experience for which we are world renowned.

Our systems and algorithms operate on one of the world's largest product catalogs and it is quite routine for our systems to operate on Terabyte scale datasets using distributed frameworks such as Apache Hadoop(Map/Reduce) and other open source technologies such as Lucene. We consistently strive to improve the customer search and browse experience.

Our team is dedicated to finding similar products in order to place contextually relevant product advertisements on each page, as well as detecting egregious cases of poor quality **data** provided by sellers. We also build algorithms and systems to discover useful relationships between the billions of items that we have in our catalog. Examples of such relationships might include items that are variation of each other and items that are accessories of other items to list a few. Inferring these valuable relationships between products allows us to deliver an effective customer search and browse experience.

Who we are looking for:

YES!

Data Warehouse Predictive Models Ordinal DBSCAN
Heirarchical Pruning Density Median Graphs Nominal
Categorical Mean Range Perceptron Cosine Similarity Standard Deviation
Dimension Summary Statistics Supervised Learning
Mode Continuous Data Feature Extraction
Self-Organizing Map Random Forests Discretization
Greedy Patterns Kohonen Feature Maps Knowledge Validation
Time Series Information Noise
Bayesian Belief Network Ensemble Methods Overfitting
Unsupervised Classification Bias Regression Analysis
Artificial Neural Network Training Clustering C4.5
Confidence Decision Trees Rule Induction Backpropogation CHAID
Pearson Correlation Sampling Support Vector Machines
Descriptions Probability CART Associations
k-Nearest Neighbors Weighting Bagging
Cross-Validation Anomaly Detection Prediction Outliers
Genetic Algorithm Visualization Boosting Proximity

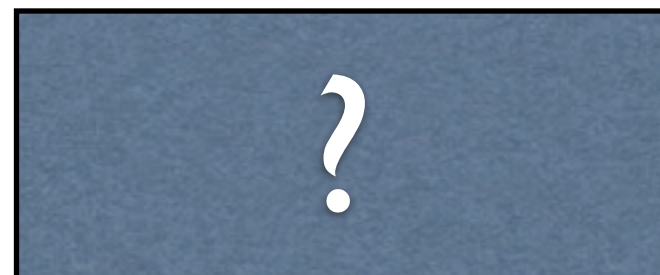
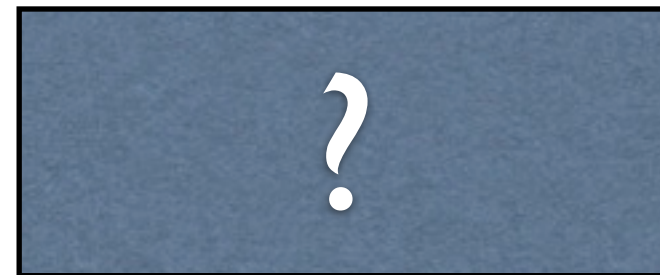


Organize Your Mind

Two general purposes



Four families of tasks



Homework

- Finish Project 1
- Reading 2