Securing Critical Infrastructures In The Financial Sector

Introduction to Predictive Analytics

Objectives

Topic	*Goal*
 *Introduction to Data Mining Process*	▪ Understand the fundamentals of the data mining process ▪ Learn about CRISP-DM
 *Prediction of security incidents using data mining*	▪ Learn about the use of data mining for predicting security events

Introduction to Data Mining Process

Introduction

- CRISP-DM process model: step-by-step approach to data mining

Primary phases:

1. **Business understanding:** Get a clear understanding of the problem, how it impacts your organization & your goals for addressing it
2. **Data understanding:** Review the data that you have, document it, & identify data management and data quality issues
3. **Data preparation:** Get your data ready to use for modeling
4. **Modeling:** Use mathematical techniques to identify patterns within your data
5. **Evaluation:** Review the patterns you have discovered and assess their potential for business use
6. **Deployment:** Put your discoveries to work in everyday business

Introduction (cont.)

- Project begins with business understanding and steps through each of the five phases of the process
- Within the cycle, there are you smaller cycles, such that several passes back and forth as you work to understand the business and the data, or to prepare data and build models
- cycle repeats while the project evaluation and experience during deployment add to the understanding of the business and inspire new projects

1. Business Understanding

- **Identifying your business goals:** find out exactly what you're trying to accomplish

- **Assessing your situation**: go deeper into fact-finding, building out a much
- fleshier explanation of the issues outlined in the business goals task

- **Defining your data-mining goals:** Break down the goals into sub-goals

- **Producing your project plan: S**pecify every step that the data miner, intend to take until the project is completed

2. Data Understanding

- **Gathering data:** verify that you have acquired the data or at least gained access to the data, tested the data access process, and verified that the data exists
- **Describing data:** source and formats of the data, the number of cases, the number and descriptions of the fields, and any other general information that may be important
- **Exploring data:** examine the data more closely. For each variable, you look at the range of values and their distributions. You'll use simple data manipulation and basic statistical techniques for further checks into the data
- **Verifying data quality:** determine whether the data quality is good enough to support your goals

3. Data Preparation

- **Selecting data:** rationale for inclusion and exclusion of data
- **Cleaning data:** make specific data corrections, excluding some cases or individual cells
- **Constructing data:** derive some new fields (for example, use the delivery date and the date when a customer placed an order to calculate how long the customer waited to receive an order), aggregate data, or otherwise create a new form of data
- **Integrating data:** merge some or all of those disparate datasets together
- **Formatting data:** convert different formats of data into most convenient format for modeling

4. Modeling

- **Selecting modeling techniques:** Narrow the list of modeling techniques based on the kinds of variables involved, the selection of techniques available in your tools, and any business considerations that are important to you
- **Designing test(s): define the** test that you'll use to determine how well your model works
- **Building model(s):** define the parameter settings, the model descriptions and the model types
- **Assessing model(s):** review the models that you've created, from a technical standpoint and also from a business standpoint

5. Evaluation

- **Evaluating results:** assess the value of your models for meeting the business goals that started the data-mining process
- **Reviewing the process:** spot issues that you might have overlooked
- **Determining the next steps:** define the recommendations for the next move

6. Deployment

- **Planning deployment : define** a strategy for putting it to work in your business
- **Planning monitoring and maintenance:** define a strategy for the ongoing review of the model's performance
- **Reporting final results:** report on the results of the project
- **Reviewing final results**: retrospect on the process to identify how it could be improved in the next cycle

Prediction of security incidents using data mining

Prediction of security incidents using data mining

- Making sense of security threats in organizations
- Use of data mining in predicting security incidents is a fast-growing trend
- Use cases on:
  - Detecting malicious activities
  - Analyzing mobile endpoints
  - Automating repetitive security tasks
  - Targeting a close zero-day vulnerabilities

Detecting malicious activity & stopping attacks

- Data mining and machine learning algorithms will help businesses to detect malicious activity faster and stop attacks before they get started

Examples:

- casino in North America: algorithms detected a data exfiltration attack that used a “connected fish tank as the entryway into the network.”
- NHS agency’s network: algorithms spotted the attack within seconds and the threat was mitigated without causing any damage to that organization

Analyzing mobile endpoints

- Use of machine by Google to analyze threats against mobile endpoints
- Seeing an opportunity to protect the growing number of bring-your-own and choose-your-own mobile devices

Example:

- MobileIron and Zimperium announced a collaboration to help enterprises adopt mobile anti-malware solutions
- MobileIron to integrate Zimperium's machine learning-based threat detection with MobileIron's security and compliance engine
- Combined solution to address challenges like detecting device, network, and application threats and immediately take automated actions to protect the company's data

Enhancing human analysis

- Data mining to helps human analysts with all aspects of the job, including detecting malicious attacks, analyzing the network, endpoint protection and vulnerability assessment

Example:

- MIT's Computer Science and Artificial Intelligence Lab (CSAIL) developed an adaptive machine learning security platform to help analysts on reviewing millions of logins each day
- The system was able to filter data and pass it onto the human analyst, reducing alerts down to around 100 per day

Automating repetitive security tasks

- Machine learning can automate repetitive tasks, enabling staff to focus on more important work

Example:

- Booz Allen Hamilton has gone down this route: using AI tools to more efficiently allocate human security resources, triaging threats so workers could focus on the most critical attacks

Targeting close zero-day vulnerabilities

- Machine learning could help close vulnerabilities, particularly zero-day threats and target largely unsecured IoT devices

Example:

- Arizona State University used machine learning to monitor traffic on the dark web to identify data relating to zero-day exploits
- Armed with this type of insight, organizations could potentially close vulnerabilities and stop patch exploits before they result in a data breach