

2017: Improve Drug Safety through Machine Learning

Add ProQuest Company Information

Pharmaceutical companies use ProQuest Dialog to support regulation compliance. Information professionals at these “Big Pharma” companies construct complicated search strategies tailored to specific drugs and run them against medical databases like Medline and Embase. The primary user need is to retrieve regulatory compliance information as quickly as possible while maintaining precision results.

Once results are retrieved from ProQuest Dialog, these results are ingested into a specialized workflow tool called the Drug Safety Triager, part of the Pi2 product suite. Each result is now called a literature reference. Each reference is manually reviewed through a process known as literature review. Highly qualified and trained screeners review each reference for a specific set of criteria. If the reference meets any of the specific criteria, it is then subjected to further review by drug safety specialists. Ultimately, references passing this second level of screening are reported to regulatory agencies such as the U.S. Food and Drug Administration (FDA) and European Medicines Agency (EMA).

Drug safety screening critical to patient safety. It is expensive due to the need for highly experienced screeners and medical practitioners. The process is time intensive due to the manual nature of the work.

The objective of this project is to leverage the power of machine learning to improve the literature review process. By using historical data generated by the manual review process, we will train an algorithm to identify references that contain reportable drug safety information. We will build a simple UI workflow suggesting the criteria identified by the algorithm to the literature screener.

This project requires students to develop an understanding of a typical drug safety review process used by top pharmaceutical companies. A successful project execution will result in a web UI that will apply the machine learning algorithm to a literature reference and make suggestions to the user.

Phase I: Minimum Functional Delivery

- Create simple machine learning algorithm
- Work with ProQuest stakeholders to define what safety criteria is a candidate for the algorithm. Identify level of difficulty for each criteria: low, medium, or high difficulty.
- Implement low difficulty criteria defined above.
- Train the algorithm with 1K-10K previous records
- Apply the machine learning algorithm to a literature reference and return possible safety criteria
- Create a REST-based web service to wrap the machine learning algorithm
- Interview Marco Ward and Bill Kivett, Product Managers for Pi2. Both can represent end users and help to prioritize user needs
- Interview Angela Duma, Manager of Operations and Analysis, Pi2. Angela runs a large team of literature screeners and can help validate the algorithm


Phase II: Successful Delivery

- Implement medium difficulty criteria defined above.
- Train the algorithm with >10,000 records
- Create a web UI literature review page that displays a reference and suggests criteria based on the results of the algorithm.

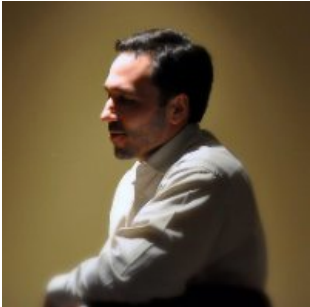
Phase III: Stretch Goal

- Implement high difficulty criteria defined above.
- Ready-to-launch UI integration with a Pi2 prototype Drug Safety Triager.

Primary Sponsor Bio

	John York
	Director of Engineering, ProQuest Dialog
	Over 15 years of experience as a software engineer, architect and technology leader. Industry experience in publishing, information, and software services. Worked for CareerSite.com, eePulse, JSTOR, and ProQuest.
	LinkedIn

Executive Sponsor Bio

	Roger Valade
	VP of Engineering, ProQuest
	Senior technology leader with extensive experience in enterprise and application architecture, software development and methodology (with an emphasis on agile), strategic planning, project and program management, offshoring in China and India, and change management. Former positions include VP, Technology for a \$200M publishing company; VP, Technical Solutions for a J2EE consultancy; and Architect at General Motors. Have managed teams of up to 105 people and budgets of nearly \$20M.
	LinkedIn

Goals

Key Skills and Project Roles

Machine Learning and Data Science	Ideal Skills
<p>Basic machine learning experience. Familiarity with algorithm creation and selection, training techniques.</p> <p>Likely majors: Computer Engineering, Data Science</p>	<p>Python/R/Java</p>
Software Development	Ideal Skills
<p>Basic programming experience. Web programming experience using HTML and CSS.</p> <p>Web service creation and core HTTP concepts</p> <p>JavaScript experience with tools like AngularJS or React.</p> <p>Likely majors: Computer Engineering, Electrical Engineering</p>	<p>Java</p> <p>HTML/CSS/Javascript</p> <p>REST</p>
Human Factors Product Design	Ideal Skills
<p>Experience presenting complex data in easy to understand way. Experience designing interactions based on structured data. Usability testing. Use case modeling, requirements gathering</p> <p>Ability to code in HTML, CSS, JavaScript a plus.</p> <p>Likely Majors: S.I. IOE)</p>	<p>Human factors UI/UX combined</p> <p>Product Interaction Evaluation</p> <p>general skills in other areas of the project Computer Science and Data Analysis</p>
Algorithm and Data Analysis	Ideal Skills
<p>Likely Majors: Mathematics, Statistics, Computer Science</p>	

Additional Skills and Knowledge Desired

- Familiarity with Pharmacology terminology
- Familiarity working with structured data
- Git version control, AWS
- Interest in product development

Section including IP/NDA Summer Internship Meetings at the Proquest Location

OTHER INFORMATION WE NEED TO COLLECT BUT WON'T BE PART OF THE STUDENT FACING PROJECT ADVERTISEMENT

Test Development Environment

2.Private, publically hosted

Students work in private GitHub utilizing their own hardware. Server environment if required is established by the sponsor.

Students will need access to a large dataset of sample search strategies that ProQuest will provide. Students will also need access to the ProQuest Dialog test search API to validate search operator weighting assumptions as well as the impacts of their algorithm. ProQuest will provide VPN accounts and access to any additional internal systems as needed.

The Documentation for the search API can be found here:

<http://docs.api.dialog.proquest.com/>