



## Master of Professional Studies in Information Science

### Sponsored Project Proposal Form – Spring 2018

Please complete the following project proposal form to sponsor an MPS Project. This form will be used to determine if your project is appropriate for MPS students and whether it is of sufficient scope for a semester long project (~400-500 person-hours). We will assign teams with complementary skills based on the skills and experience you list in this form. We will also share most of this form with the students to help them make their top project choices before we assign the projects.

Please direct any questions to the MPS Project Coordinator: [is-mps-projects@cornell.edu](mailto:is-mps-projects@cornell.edu)

<b>Sponsor Name</b>				<b>Date</b>	
<b>Contact Name(s)</b>		<b>Email(s)</b>		<b>Phone</b>	

#### Description of the Sponsor

At Red Hat, we connect an innovative community of customers, partners, and contributors to deliver an open source stack of trusted, high-performing solutions. We offer cloud, Linux, middleware, storage, and virtualization technologies, together with award-winning global customer support, consulting, and implementation services. Red Hat is a rapidly growing company supporting more than 90% of Fortune 500 companies.

The AI Center of Excellence is a group within the CTO Office of Red Hat. Our mission is to define and work on the vision of AI within Red Hat. This means enhancing internal tooling and processes as well as influencing customer facing services and products. We work with a lot of teams within Red Hat, enabling them in the context of AI and deliver Proof of Concepts (POC) that have a real impact on products and services used by customers.

#### Please indicate which academic year and semester you would like to propose your project.

<b>Year</b>	2018	<b>Semester</b>	<input type="checkbox"/> Fall	<input checked="" type="checkbox"/> Spring	
-------------	------	-----------------	-------------------------------	--	--

#### Project Title

AI enhanced tooling for developers of an IT analysis and problem remediation service

#### Project Goal or Description

##### Deliverable:

A web application from which a user can select a combination of features, and submit the selection for analysis. The application can perform different variations of correlation analysis for the selection, and present the user with the results of the analysis. The data used will be data that has been collected by Red Hat Insights.

#### Terminology:

- Parser
  - Component of Red Hat Insights which analyses system logs and outputs structured data
    - A feature is equivalent to a row in the parser output data
- Feature
  - Singular categorical or numeric attribute which is used in conjunction with other features (correlation) to validate rule components
- Rule
  - A feature or multiple features which identifies a system configuration with suspected issues that requires reconfiguration or action by an administrator.

#### Background:

Red Hat Insights is a hosted service that helps with identifying, predicting and preventing problems in a customer's IT environment before they occur. The process starts by collecting data from running systems; like installed packages, configurations and log files. This data is sent to Red Hat for analysis. The Red Hat Insights service consumes this data, identifies problems, and offers a solution to fix the identified issues.

For example, you could have a vulnerability related to your remote login server SSH that allows for a password guessing attack. Given this vulnerability is only relevant if a specific configuration flag is set, Red Hat Insights would run a "rule" to detect this vulnerability based on the system data which was sent to Red Hat. Red Hat Insights then offers a script the customer can run to fix the configuration issue.

This is a typical rule being developed in Python by rules developers. For their work they not only need to know about possible problems and how to fix them, but also about the likelihood of configurations and interdependencies of systems. We will focus on the latter and provide the developers with a tailored web based tooling to explore the data and find correlations.

#### **Project Details:**

- The web UI will have a search function to make all features searchable by name or keyword
- Users of the UI will search for one to many features, and have the choice to add each feature into the "shopping cart"
- Once the user of the UI has added all desired features to the "shopping cart", the user will submit the contents for correlation analysis
- Different variations of correlation analysis will be executed depending upon the types of the features selected (e.g. Categorical, numerical), the distribution of the data for those features, and other factors.
- Once the analysis is complete, the user will be presented with a correlation table in the UI for the submitted features.
  - The user will also be presented with correlation tables for each combination of the selected features with the highest feature correlation score listed first.
- Bonus:
  - The user will be presented with additional information regarding the features selected such as outliers of a single feature or a combination of features.

#### **Project Roles:**

- Data Analysis
  - Automate analysis and reporting on parsers
    - Distribution
    - Correlation
    - Visualization: e.g. box plot
  - Perform manual analysis of parsers which yielded incomplete results as part of the automated analysis above
- UI
  - Developing the web UI following a UX spec provided by Red Hat
  - Developing the backend services to support the web UI
  - Deploying the web UI in OpenShift for use by users

#### **What activities are necessary to achieve the project goal?**

In the Data Analysis role, you will assist our data scientist to explore the data, find correlation examples and visualize the results. For example you will import data of installed packages on system into a Python pandas dataframe and use scikit-learn tools to find correlations of packages. This should be documented and displayed in a Jupyter notebook and the results presented and discussed with the team.

The UI role will work with our UX team to package the data analysis workflow into a web UI and make it accessible for users to select features for analysis, submit them for analysis, and view the results of the analysis. This includes the backend services required to pull data for analysis and execute the analysis code provided by the Data Analysis team. Finally, the results must be presented back to the user via the UI. The UI itself will need to be developed and ideally, should be based on the PatternFly library (bootstrap) for interface components, styling, and graphs..

The entire project will be done in a agile way with a short feedback cycle and close collaboration between the rules developers, UI/UX developer, data scientists, and the actual users.

**What outcome would determine that the project is a success? Do you expect specific deliverables?**

This is not an academic project to proof some correlations.

The goal is to apply basic and modern machine learning approaches to have an actual impact on the workflow of rules developers by making them more efficient and accurate in their rule development. This does not necessarily mean a polished and full featured product - but a usable proof of concept to demonstrate improved rule development leveraging a modern UI to guide rule developers in selecting features for analysis and results that are informed by machine learning algorithms.

If we can provide and prove this, the project is a success.

**What are the skills and experience must the students already know to start work on the project?**

Please be specific and keep in mind that students will be building their skills during the duration of the project.

**Data Science**

- machine learning tools such as python, scikit-learn, and jupyter notebooks or similar.
- understanding of basic machine learning models, anomaly detection
- python development experience

**UX/UI**

- python for backend services work
- CSS and javascript, preferably Angular experience

**What are the skills and experience required to complete the project that the students may learn while completing the project?**

Besides the obvious practically oriented data science and UX/UI part, you will learn how to collaborate in a globally distributed open source team. We have people from US, China and Europe involved. Furthermore, you will gain experience deploying applications in an enterprise-like manner.

Most of the development will happen on github and virtual meetings.

**The project representative must be available 30 minutes per week for status reports, the interim report, and the final presentation. As the project sponsor, are you able to make this time commitment?**

Yes. Please elaborate.

Yes, you will closely collaborate with stakeholders and developers within Red Hat. Those will work on the same project and contribute actual work. It'll be a small, but cross functional team. So everybody will know and work with everybody on a regular basis.

**Some sponsors may choose to spend additional time with the student teams, e.g. phone contacts for monthly status discussions, reviewing research results, providing midpoint project feedback, and offering input to the final deliverables in advance of its completion. As the project sponsor, are you available to participate in these or any additional activities?**

Yes. Please elaborate.

see above

**The project representative needs to facilitate access to company resources as needed and approve expenses. As the project sponsor, are you able to facilitate access to such resources, should the need come up?**

Yes. Please elaborate.

Hugh? Party? I'm all in 

Steven Huels will be the contact for these items

**Please consider other contributions listed below. Are you willing to make these contributions? (check all that apply)**

- Provide existing industry and company data as background at the beginning of the project.  
 Pay one or more team members to travel to your location for initial briefing / work session / final presentation.

Please elaborate.

We will give an introductory session on the product Red Hat Insights and its position in the market.

Please send your completed project proposal to the MPS Project Coordinator: [is-mps-projects@cornell.edu](mailto:is-mps-projects@cornell.edu)