Stack Watson

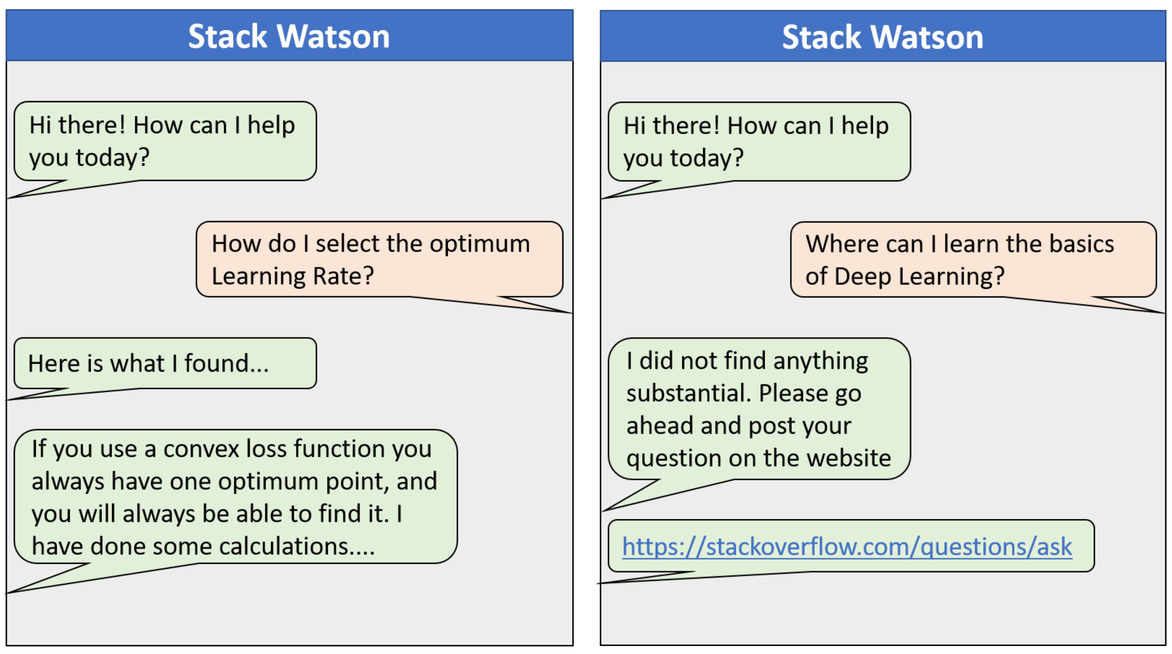
The Friendly S.O. Bot

Smit Mehta, Xue (Calvin) Cao | Advisor: Prof. Chris Asakiewicz

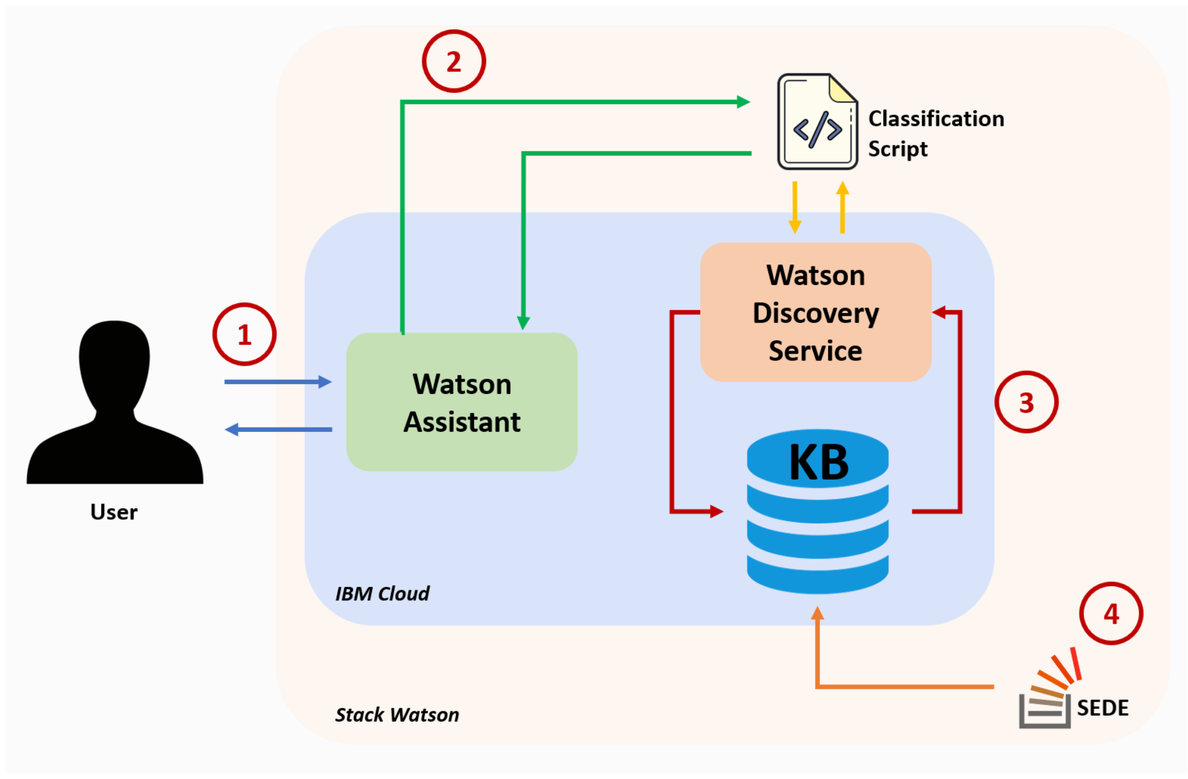
STEVENS INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®
Business Intelligence & Analytics

Stack Overflow Bot

- Provide real-time help to programmers on trivial programming issues using the **wealth of information** already existing on the Stack Overflow website
- This will also help in reducing the workload on administrators that have to monitor incoming questions for **duplicates**



Architecture



1. The interface between a user and Stack Watson, facilitated by Watson Assistant.
   - User will ask ***"natural language queries"*** which WA will pass to the AI agent
   - Watson Assistant will present the answer with the highest confidence level
   - If an answer is not found, it will prompt the user to post it on SO website
2. **Topic Modeling** is used to categorize the question into different tags to ensure relevance
3. Watson Discovery Service interacts with the **Knowledge Base *("KB")*** to return relevant answers. The threshold for relevancy can be set by us depending on the level of training provided to the application
4. KB is regularly updated with new questions being added to the SO website through automated extraction from the **SEDE data dumps**

Data & Scope

- ***"Stack Exchange Data Dumps"*** by Stack Exchange, Inc. via archive.org; specifically the Data Science Stack Exchange
- Due to the high volume of data and limited resources, we will limit the scope of this project to a particular topic area
- The data consists of all the questions with the following tags: ***<machine-learning>*** and ***<neural-network>***

Impact & Future Scope

- When someone posts a question on Stack Overflow, they have to wait for some time before they can get an answer *(sometimes it's even days!)*
- A chatbot for trivial questions would **eliminate the time lag** and make the programmer more **productive**
- This application would also **save the subject matter expert time** and as a result focus on more pressing and important matters

Future Scope:

- **Scale it up** with more training data and also include other stack exchange websites content
- Adding additional features such as **checking the quality of the questions** being asked (another time saving option for mods)
- **Return links to additional info and sample code** *(if available)* by connecting to a central repository