

# Machine Learning to Support Big Data System Acquisition

## Problem:

In rapidly evolving technology domains, there is no efficient way to create decision support tools with up-to-date, comprehensive product information for comparison, evaluation, and selection.

- Choose a modern technology stack (playbook.cio.gov)
- The subject matter competencies for successful enterprise IT system acquisition are often missing in government. (GAO)

## Solution Approach

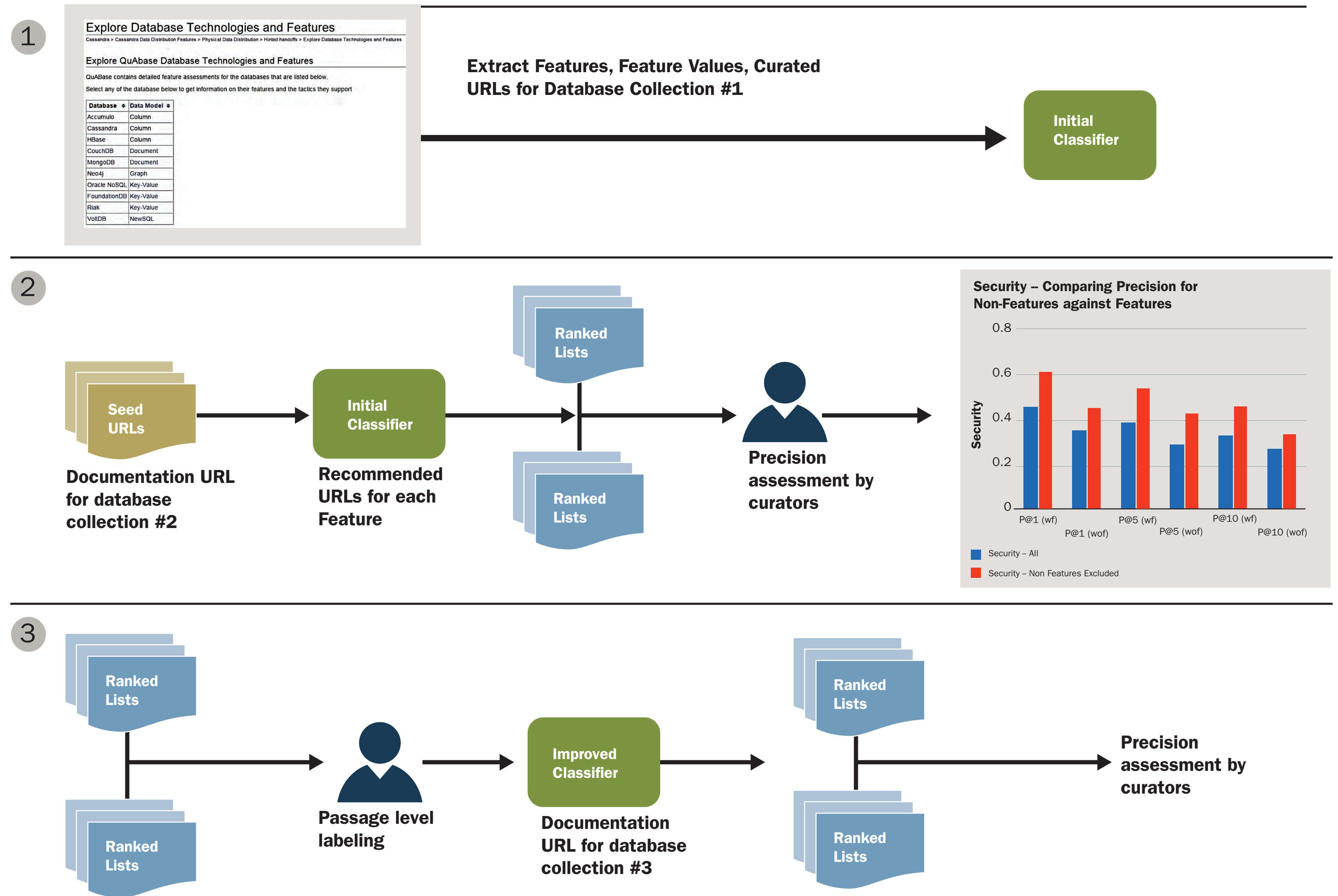
- Use machine learning to populate knowledge bases for decision support with latest information about rapidly evolving technology domains.
- Validate approach with big data technologies to show feasibility for broader application (e.g. SOA, analytics) in acquisition and IT system modernization

## Results Towards Automating Curation

- Demonstrated ability to classify documents according to knowledge base feature taxonomy for positive feature values
- Classifier precision improved as training set was augmented

### Research Team Leads

Ian Gorton, PhD, Prof. Yiming Yang, CMU LTI



Accuracy of Curating Consistency and Security for 4 Different Distributed Databases.

Product Name	Consistency Feature Model	Security	Overall
Oracle	90.10	70.80	<b>80.40</b>
Berkeley	81.80	83.30	<b>82.50</b>
Couchbase	66.70	79.20	<b>71.90</b>
Redis	84.80	91.70	<b>87.70</b>
Mean	80.16	81.25	<b>80.65</b>

Contact: John Klein jklein@sei.cmu.edu

Copyright 2015 Carnegie Mellon University

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Department of Defense.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

This material has been approved for public release and unlimited distribution except as restricted below.

Internal use:\* Permission to reproduce this material and to prepare derivative works from this material for internal use is granted, provided the copyright and "No Warranty" statements are included with all reproductions and derivative works.

External use:\* This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other external and/or commercial use. Requests for permission should be directed to the Software Engineering Institute at [permission@sei.cmu.edu](mailto:permission@sei.cmu.edu).

\* These restrictions do not apply to U.S. government entities.

Carnegie Mellon® is registered in the U.S. Patent and Trademark Office by Carnegie Mellon University.

DM-0002948