CS 5542 BIG DATA ANALYTICS AND APPLICATIONS

KANSAS CITY FOUNTAIN MONUMENT IDENTIFIER – REPORT #1

TEAM #2 AJNN: #4 JACQULYN BATSON | #20 CHIA-HUI AMY LIN | #36 NICKSON SIGEI | #46 NANCY YANG

❖ Project Objectives

▪ Significance

- This app is unique in that it will be able to tell the artist and year of work when a picture of building is taken. Apps that exist that are related to this are not able to capture a photo and give such information.

▪ Features: Use Case/Scenario

- The app will be able to take a photo and/or when a photo is chosen from the gallery will predict and tell which Fountain monument it is.

❖ Approach

Machine learning approach will be used whereby pictures will be put into categories, trained and using the chosen tool will predict which Fountain Monument it belongs to.

▪ Data Sources

- Download from google image. Following is the link of the data source: https://www.dropbox.com/home/Public/BigDataAppProjPic
- Amazon Web Services public datasets http://aws.amazon.com/datasets
- Freebase http://www.freebase.com/
- UCI Machine Learning Repository

▪ Analytic Tool

- Excel sheets. Results from the error rate from the program.
- Spark RDD.
- Others to be decided and investigate.

▪ Analytical Tasks

- Error rate comparison with lab tutorial results.
- Changing the training dataset VS testing dataset ratio, compare the error rate result.
- Changing the number of training dataset number, keeping testing dataset number fixed, compare the result
- Changing the number of testing dataset number, keeping the training dataset number fixed, compare the result.
- Others to be developed.

- Expected Inputs/Outputs
  - Input: 3-4 categories of individual famous fountains in Kansas City (3-4 different fountains). Each fountain will have around 25-30 training dataset and 10-15 picture of testing dataset. Since our tasks is to recognize different fountains, we need to distinguish fountains.

    Example Input →

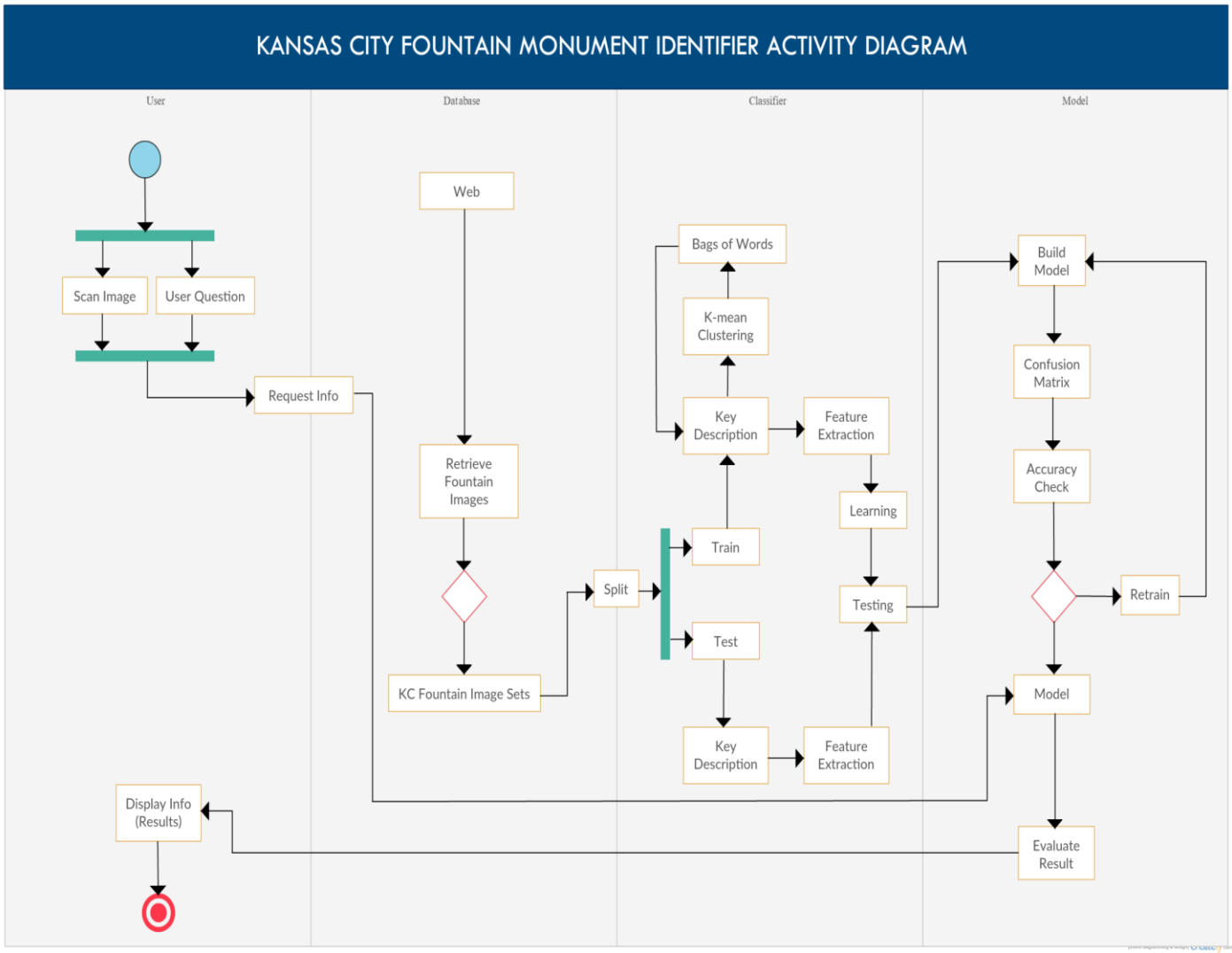
  - Output: The fountain name.

    Example Output → *J.C. Nichols Memorial Fountain*, by Henri-Léon Gréber

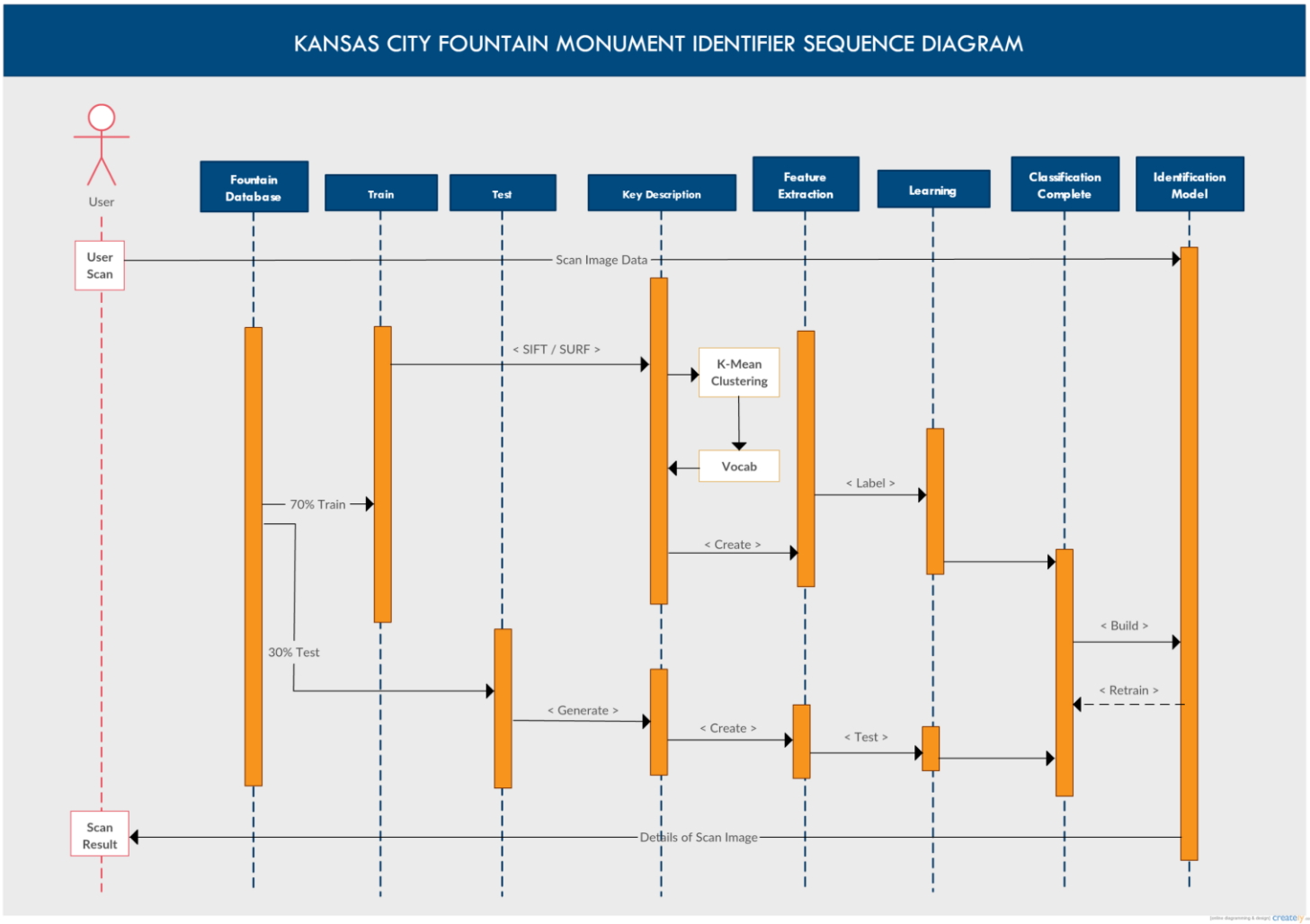
❖ Application Specification

- System Specification (Big Data Analytics Server/Client)
  - ✓ Features, workflow, technologies

- Activity Diagram(workflow, data, task)



- Sequence Diagram(interaction/collaboration)



- Feature Specification
  - Display real-time information/details of certain fountain monument in Kansas City.

- Operation Specification(Input/Output, exceptions)
  - <u>Input:</u>
    Scan image of monuments in Kansas City from user's device or allow user to ask questions using the device when it has the image or pointing toward the image.

  - <u>Output:</u>
    Detail information of the monument that user requests. (Name of the monument, Establish Year, Name of the Artist, Background details … etc.) Show/pop-up details in real-time when the user is pointing toward the image or when the user asks the question of specific monument.

  - <u>Exceptions:</u>
    Monuments outside of Kansas City might not be identified.

- Existing Applications/Services Used

  ✓ Google Arts & Culture
  https://www.google.com/culturalinstitute/beta/

  - Description:
    Partner with different institutions and museums around the world. An app to discover multiple art collections and stories on any device (smartphones, tablets, laptops…). Provide the user with an easy way to explore cultural treasures in details like you're personally visiting/experiencing the piece of art. Have an experiments webpage for developers and programmers to combine technology with artworks. Techniques include machine learning, image classification, serendipity, image similarity, real-time 3D, data visualization, meta data and image similarity.

❖ Implementation

- An Android App that takes/scans photos with camera from the device.
- Use Spark to identify the image with the trained/tested datasets.
- Return details/info of the images on user interface.

❖ Project Management

- Plan & Project Timelines

  - 03/11/2017 – Set up UI + Train/Test Data. Improve accuracy as needed.
  - 03/18/2017 – Build modules.
  - 03/25/2017 – Resolve any potential/upcoming issues for each modules.
  - 04/04/2017 – Android App/platform development.
  - 04/11/2017 – Finish up app + Connect all modules.
  - 05/04/2017 – Testing + wrapping up project.

- Implementation Status Report

  ✓ Work Completed
  - Description: Collect fountain data, diagrams(workflow and what model to build), narrow down the whole structure for the project.

  - Responsibility (Task – Person):
    Fountain Data collection – Nancy
    Diagrams/workflow/collaboration – Amy
    Code contribution/Spark – Jackie
    Information collect – Nick

- Time taken (hours): 10 hours
- Contributions(members %): Jackie 26% | Nancy 26% | Amy 26% | Nick 22%

✓ Work to be Completed
- Description: UI, set up three modules, train/test data, develop application, connect modules with app.
- Responsibility: We'll all be working on the same part together one at a time and push through the process.
- Time to be taken (estimated hours): 30+ hours (Speed up process so we can have time to test/deal with final stage of the project.)

✓ Issues/Concerns
- Precision/accuracy for the identification outcome.
- Verbal interaction part from user to the application.