Capstone: Mushroom Classification

Western Governors University

002957010

Table of Contents

[**Prompt A** 4](#_Toc25684)

[Letter of Transmittal 4](#_Toc25685)

[Project Recommendation 5](#_Toc25686)

[Problem Summary 5](#_Toc25687)

[Application Benefits 5](#_Toc25688)

[Application Description 5](#_Toc25689)

[Data Description 5](#_Toc25690)

[Objective and Hypothesis 6](#_Toc25691)

[Methodology 7](#_Toc25692)

[Funding Requirements 7](#_Toc25693)

[Stakeholders Impact 7](#_Toc25694)

[Data Precautions 8](#_Toc25695)

[Developer Expertise 8](#_Toc25696)

[**Prompt B** 9](#_Toc25697)

[Project Proposal 9](#_Toc25698)

[PROBLEM STATEMENT 9](#_Toc25699)

[CUSTOMER SUMMARY 9](#_Toc25700)

[EXISTING SYSTEM ANALYSIS 9](#_Toc25701)

[DATA 9](#_Toc25702)

[PROJECT METHODOLOGY 10](#_Toc25703)

[PROJECT OUTCOMES 11](#_Toc25704)

[IMPLEMENTATION PLAN 11](#_Toc25705)

[EVALUATION PLAN 12](#_Toc25706)

[RESOURCES AND COSTS 12](#_Toc25707)

[TIMELINE AND MILESTONES 13](#_Toc25708)

[**Prompt C** 14](#_Toc25709)

[Application Files 14](#_Toc25710)

[**Prompt D** 15](#_Toc25711)

[Post-implementation Report 15](#_Toc25712)

[Project purpose 15](#_Toc25713)

[Datasets 15](#_Toc25714)

[Data product code 16](#_Toc25715)

[Hypothesis verification 16](#_Toc25716)

[Effective visualizations and reporting 17](#_Toc25717)

[Accuracy analysis 17](#_Toc25718)

[Application testing 20](#_Toc25719)

[**Appendices** 20](#_Toc25720)

[Installation Guide 21](#_Toc25721)

[User Guide 22](#_Toc25722)

[Summation of Learning Experience 23](#_Toc25723)

[References 24](#_Toc25724)

# Prompt A

## Letter of Transmittal

Steven Greensweight  
5525 Parkview Hills Ln  
Fort Worth, TX, 76179

Zolka Tessana,   
President  
Mushroom Investigation Group  
176 Spell Ln  
Ythrn, Icewind Dale, 88088

Dear Ms. Zolka Tessana,

A representative of your group, Ella Vator, reached out to me requesting a tool to assist your operations by streamlining the identifying process for new mushrooms that you come across. It was explained that accuracy and accessibility were of the utmost importance due to the nature of your business. I am proposing the creation of a classification program that will allow members of your group to enter the attributes of a mushroom they find and determine whether it is edible or poisonous.

Creating an accurate and easy to use tool will enable your employees to identify the edibility of foraged mushrooms quickly and safely. This will both raise employee morale and reduce the number of insurance-claims from people eating poisonous mushrooms. Both will prove to be a cost-savings, which allows for additional budget to be allocated elsewhere.

The development of this project is estimated to be $15, 000 and will cover development costs, labor, six months of support, and any additional hardware that is needed. My five years of experience in project management, machine learning, and the python programming language will allow me to keep costs down while maintaining the scope outlined in the below document.

If you have any questions about the project or the below document, please reach out to me at (940)859-4096. I look forward to our partnership.

Sincerely,

Steven Greensweight

## Project Recommendation

### Problem Summary

Mushroom Investigation Group is a regional organization based in the foraging and selling of mushrooms to local populations. To control the rapidly growing cost of insurance claims due to the consumption of poisonous mushrooms, they need a way to classify mushrooms quickly and accurately.

Current methods to classify mushrooms range include running various chemical tests in a laboratory setting, matching mushrooms to a dictionary of known mushrooms, human-testing, hiring certified mushroom professionals, training certified mushroom professionals. Each of these methodologies increase the time and cost of day-to-day operations.

### Application Benefits

The Mushroom Classification program will provide a streamlined method of identifying whether a mushroom is poisonous or edible. The program will allow the user to select mushroom attributes that correspond to a mushroom they have found in the wild and receive a quick prediction about the edibility of the mushroom. Out of the mentioned options, this will decrease both time and costs associated with identifying foraged fungus. Other third-party solutions are costly and require intensive training to use. The proposed solution is cost-effective and lightweight with an easy-to-use interface.

### Application Description

The application will consist of a user-interface that allows the user to select mushroom attributes from dropdown menus that correspond to the mushroom they are identifying. It will then test it against a trained model and return a result of poisonous or edible. All done with no risk of poison to the user

### Data Description

The initial data used to train our machine learning model will include all attributes found in the table below. As operations continue, users will submit their recorded data which will be combined with the current data and our model will continue to be retrained for improved accuracy.

Mushroom Attributes

|  |  |
| --- | --- |
| class | edible=e, poisonous=p |
| cap-shape | bell=b, conical=c, convex=x, flat=f, knobbed=k, sunken=s |
| cap-surface | fibrous=f, grooves=g, scaly=y, smooth=s |
| cap-color | brown=n, buff=b, cinnamon=c, gray=g, green=r, pink=p, purple=u, red=e, white=w, yellow=y |
| bruises | bruises=t, no=f |
| odor | almond=a, anise=l, creosote=c, fishy=y, foul=f, musty=m, none=n, pungent=p, spicy=s |
| gill-attachment | attached=a, descending=d, free=f, notched=n |
| gill-spacing | close=c, crowded=w, distant=d |
| gill-size | broad=b, narrow=n |
| gill-color | black=k, brown=n, buff=b, chocolate=h, gray=g, green=r, orange=o, pink=p, purple=u, red=e, white=w, yellow=y |
| stalk-shape | enlarging=e, tapering=t |
| stalk-root | bulbous=b, club=c, cup=u, equal=e, rhizomorphs=z, rooted=r, missing=? |
| stalk-surface-above-ring | fibrous=f, scaly=y, silky=k, smooth=s |
| stalk-surface-below-ring | fibrous=f, scaly=y, silky=k, smooth=s |
| stalk-color-above-ring | brown=n, buff=b, cinnamon=c, gray=g, orange=o, pink=p, red=e, white=w, yellow=y |
| stalk-color-below-ring | brown=n, buff=b, cinnamon=c, gray=g, orange=o, pink=p, red=e, white=w, yellow=y |
| veil-type | partial=p, universal=u |
| veil-color | brown=n, orange=o, white=w, yellow=y |
| ring-number | none=n, one=o, two=t |
| ring-type | cobwebby=c, evanescent=e, flaring=f, large=l, none=n, pendant=p, sheathing=s, zone=z |
| spore-print-color | black=k, brown=n, buff=b, chocolate=h, green=r, orange=o, purple=u, white=w, yellow=y |
| population | abundant=a, clustered=c, numerous=n, scattered=s, several=v, solitary=y |
| habitat | grasses=g, leaves=l, meadows=m, paths=p, urban=u, waste=w, woods=d |

### Objective and Hypothesis

A mushroom can be classified as edible or poisonous based on its attributes. These attribute combinations can be used to train a machine learning model to be able to identify mushrooms in this way.

The main objective of this project is to reduce the risk of identifying edible mushrooms. Employees of the MIG will be able to forage freely without fear of a fungal faux pas. The secondary objective is to cut back on the time and cost associated with identifying the edibility of mushrooms found in the wild.

### Methodology

The methodology that will be used for the development cycle will be the traditional waterfall project management methodology. It is the most cost-effective option for small projects that have fixed requirements. This project meets both of those criteria.

The waterfall methodology follows five stages (ionos.com 2019):

1. Analysis – Gather the requirements of stakeholders and identify project scope
2. Design – Create a mockup that meets all requirements and fits in the scope
3. Implementation – Develop the application to the design specifications
4. Testing – Integrate software into the environment and test functionality
5. Maintenance – Release for production use and monitor function.

### Funding Requirements

The costs for the project are totaled at $25,000 broken down into the following:

* Development - $20,000
* Support for six months - $4,000
* Hardware - $1,000

Any additional support after the first six months of production-use will be billed at $150/hour.

### Stakeholders Impact

The program will allow employees to identify edible mushrooms quickly and safely. This will improve the safety of the end-user and cut back on hospital visits due to mushroom related poisonings. Because of the decreased medical necessity, insurance premiums will also decrease, saving money for both the MIG and its employees.

The improved workflow will also allow for more foraging to be done during business hours, allowing more time for additional projects with increased budgetary room.

### Data Precautions

The initial data being used is under no regulations by the local or national government. All data gathered by employees of the MIG will be proprietary information and will need to be kept in a secured server hosted locally at the MIG headquarters in Ythrin, Icewind Dale.

The mushroom data will be the only data stored on the server and is not subject to any regulatory frameworks. All sales data will be kept separate.

### Developer Expertise

Our development team has a combined twenty-seven years of experience in software development, technical service, machine learning, and customer support. Our project management experience is supported by the ITILv4 Foundation certification and the CompTIA Project+ certification.

Our team has also been delivering projects completed within the designated timeframe and scope for the past five years. Each developer has academic and professional experience and achievements that provide a strong foundation for the completion of this project.

# Prompt B

## Project Proposal

### PROBLEM STATEMENT

Mushroom related poisonings cost the MIG hundreds of thousands of dollars each year. Around 7,500 poisonous mushrooms ingestions are reported to poison control centers across the United States each year (Gold et al., 2021). Not only does it cost the company monetarily, it also severely damages employee morale whenever a peer is hospitalized due to working conditions.

As operations grow, the costs of mushroom incidents, both monetarily and emotionally, will continue to increase.

### CUSTOMER SUMMARY

The customer of this program is the Mushroom Investigation Group (MIG). The MIG’s focus is foraging wild mushrooms to sell to the local populations and perform research. All employees of the foraging division need to be able to identify whether a mushroom is poisonous or edible.

### EXISTING SYSTEM ANALYSIS

The MIG is currently using two methods for mushroom classification:

1. Chemical lab testing – This is a major operations cost and slows down the company.
2. Taste-testing – This is a liability to the company and is the number one cause of employee hospitalization.

### DATA

The initial data used to train our machine learning model will include all attributes found in the table below. As operations continue, users will submit their recorded data which will be combined with the current data and our model will continue to be retrained for improved accuracy.

Mushroom Attributes

|  |  |
| --- | --- |
| class | edible=e, poisonous=p |
| cap-shape | bell=b, conical=c, convex=x, flat=f, knobbed=k, sunken=s |
| cap-surface | fibrous=f, grooves=g, scaly=y, smooth=s |
| cap-color | brown=n, buff=b, cinnamon=c, gray=g, green=r, pink=p, purple=u, red=e, white=w, yellow=y |
| bruises | bruises=t, no=f |
| odor | almond=a, anise=l, creosote=c, fishy=y, foul=f, musty=m, none=n, pungent=p, spicy=s |
| gill-attachment | attached=a, descending=d, free=f, notched=n |
| gill-spacing | close=c, crowded=w, distant=d |
| gill-size | broad=b, narrow=n |
| gill-color | black=k, brown=n, buff=b, chocolate=h, gray=g, green=r, orange=o, pink=p, purple=u, red=e, white=w, yellow=y |
| stalk-shape | enlarging=e, tapering=t |
| stalk-root | bulbous=b, club=c, cup=u, equal=e, rhizomorphs=z, rooted=r, missing=? |
| stalk-surface-above-ring | fibrous=f, scaly=y, silky=k, smooth=s |
| stalk-surface-below-ring | fibrous=f, scaly=y, silky=k, smooth=s |
| stalk-color-above-ring | brown=n, buff=b, cinnamon=c, gray=g, orange=o, pink=p, red=e, white=w, yellow=y |
| stalk-color-below-ring | brown=n, buff=b, cinnamon=c, gray=g, orange=o, pink=p, red=e, white=w, yellow=y |
| veil-type | partial=p, universal=u |
| veil-color | brown=n, orange=o, white=w, yellow=y |
| ring-number | none=n, one=o, two=t |
| ring-type | cobwebby=c, evanescent=e, flaring=f, large=l, none=n, pendant=p, sheathing=s, zone=z |
| spore-print-color | black=k, brown=n, buff=b, chocolate=h, green=r, orange=o, purple=u, white=w, yellow=y |
| population | abundant=a, clustered=c, numerous=n, scattered=s, several=v, solitary=y |
| habitat | grasses=g, leaves=l, meadows=m, paths=p, urban=u, waste=w, woods=d |

### PROJECT METHODOLOGY

The methodology that will be used for the development cycle will be the traditional waterfall project management methodology. It is the most cost-effective option for small projects that have fixed requirements. This project meets both of those criteria.

The waterfall methodology follows five stages (ionos.com 2019):

1. Analysis – Gather the requirements of stakeholders and identify project scope
2. Design – Create a mockup that meets all requirements and fits in the scope. This includes the user interface diagram, program architecture, hardware architecture, etc.
3. Implementation – Develop the application to the design specifications. Ensure that the development follows the design documents as closely as possible.
4. Testing – Integrate software into the environment and test functionality. Perform unit tests when implementing new features and regression tests when fixing current features.
5. Maintenance – Release for production use and monitor function. Patch any bugs that are discovered throughout the life of the software. Add additional features that are necessary for production.

### PROJECT OUTCOMES

Upon completion, the finished program will be delivered to the MIG. The program will include a locally hosted database of mushroom information. Users will access the application through a web portal that will be administrated by the MIG IT team.

Training will be performed for the month after deployment of the program, and additional support is covered for five months after that.

Documentation will also be provided to the customer. This documentation will cover the implementation, specifications, and workflow expected for the mushroom classification program.

### IMPLEMENTATION PLAN

1. Obtain balanced data that contains a similar number of edible and poisonous mushrooms with varying attributes.
2. Tune the data by removing unnecessary or null attributes and converting to datatypes that can be used by the machine learning model.
3. Develop the model using a Support Vector Machine algorithm that will be used to classify the mushrooms into edible and poisonous.
4. Return to steps 2 and 3 as needed to fine-tune the data and model.
5. Create the user interface that will accept mushroom attributes
6. Integrate the program into the MIG environment using a new server running Ubuntu version 20.04 that will only host mushroom data and the application.
7. Begin user acceptance testing. The program should be a net positive impact on workflow.
8. Finalize product and ensure stakeholder needs have been met

### EVALUATION PLAN

Evaluation will continue to be performed throughout the development of the mushroom classification program (Synoptek, 2022).

1. Unit tests will be done to ensure the initial model is making predictions within a 95% accuracy.
2. One a model has been selected component testing will be performed.
3. Integration testing will be performed to ensure that the program fits into the current workflow of the employees. This also validates that all current systems will work with the product.
4. The final evaluation is user acceptance testing. This stage will consist of employees testing the program and giving determining if it assists or negatively impacts their work.

After all evaluation has been complete, the project will be considered at its end.

### RESOURCES AND COSTS

|  |  |
| --- | --- |
| **Resource** | **Cost** |
| 2x Developer for 120 hours each | $20,000 |
| HP Proliant DL360p Server | $1,000 |
| Support for six months | $4,000 |
| **Total:** | $25,000 |

### TIMELINE AND MILESTONES

|  |  |  |  |
| --- | --- | --- | --- |
| ***Phase*** | ***Task*** | ***Start*** | ***End*** |
| *Analysis* | Gather Stakeholder requirements | 1/2/2023 | 01/3/2023 |
|  | Identify project scope | 1/4/2023 | 01/6/2023 |
| *Design* | Design framework for program | 1/9/2023 | 1/11/2023 |
|  | Design hardware infrastructure | 1/11/2023 | 1/12/2023 |
|  | Design Database structure | 1/12/2023 | 1/13/2023 |
| *Implementation* | Setup database | 1/16/2023 | 1/16/2023 |
|  | Install hardware infrastructure | 1/17/2023 | 1/17/2023 |
|  | Execute on software design | 1/18/2023 | 2/8/2023 |
| *Testing* | Unit testing | 2/9/2023 | 2/10/2023 |
|  | User beta testing | 2/13/2023 | 2/17/2023 |
|  | User acceptance testing | 2/20/2023 | 2/28/2023 |
| *Maintenance* | Provide support and training | 3/1/2023 | 8/31/2023 |

# Prompt C

## Application Files

\C964

\Data Descriptions.txt A dictionary of the mushroom attributes

\Mushroom Classification.ipynb contains the application

\mushrooms.csv initial data set for training

\README.md readme

# Prompt D

## Post-implementation Report

### Project purpose

The mushroom classification project provides the ability to predict whether a mushroom will be edible or poisonous. It does so by comparing the weights of different mushroom attributes in respect to the occurrences of edible or poisonous classifications. It then places the mushroom into a category using a support vector machine model and classifies it as edible or poisonous with a high degree of confidence.

### Datasets

The data we used to train our model is sourced from [Kaggle](https://www.kaggle.com/datasets/uciml/mushroom-classification). As operations continue, users will submit data they have found from their foraging. The data includes all attributes found in this table:

Mushroom Attributes

|  |  |
| --- | --- |
| class | edible=e, poisonous=p |
| cap-shape | bell=b, conical=c, convex=x, flat=f, knobbed=k, sunken=s |
| cap-surface | fibrous=f, grooves=g, scaly=y, smooth=s |
| cap-color | brown=n, buff=b, cinnamon=c, gray=g, green=r, pink=p, purple=u, red=e, white=w, yellow=y |
| bruises | bruises=t, no=f |
| odor | almond=a, anise=l, creosote=c, fishy=y, foul=f, musty=m, none=n, pungent=p, spicy=s |
| gill-attachment | attached=a, descending=d, free=f, notched=n |
| gill-spacing | close=c, crowded=w, distant=d |
| gill-size | broad=b, narrow=n |
| gill-color | black=k, brown=n, buff=b, chocolate=h, gray=g, green=r, orange=o, pink=p, purple=u, red=e, white=w, yellow=y |
| stalk-shape | enlarging=e, tapering=t |
| stalk-root | bulbous=b, club=c, cup=u, equal=e, rhizomorphs=z, rooted=r, missing=? |
| stalk-surface-above-ring | fibrous=f, scaly=y, silky=k, smooth=s |
| stalk-surface-below-ring | fibrous=f, scaly=y, silky=k, smooth=s |
| stalk-color-above-ring | brown=n, buff=b, cinnamon=c, gray=g, orange=o, pink=p, red=e, white=w, yellow=y |
| stalk-color-below-ring | brown=n, buff=b, cinnamon=c, gray=g, orange=o, pink=p, red=e, white=w, yellow=y |
| veil-type | partial=p, universal=u |
| veil-color | brown=n, orange=o, white=w, yellow=y |
| ring-number | none=n, one=o, two=t |
| ring-type | cobwebby=c, evanescent=e, flaring=f, large=l, none=n, pendant=p, sheathing=s, zone=z |
| spore-print-color | black=k, brown=n, buff=b, chocolate=h, green=r, orange=o, purple=u, white=w, yellow=y |
| population | abundant=a, clustered=c, numerous=n, scattered=s, several=v, solitary=y |
| habitat | grasses=g, leaves=l, meadows=m, paths=p, urban=u, waste=w, woods=d |

### Data product code

The user will have access to an interface that includes a series of dropdowns each corresponding to a mushroom attribute. When the user clicks on the “Predict” button it will convert the data given to integers and push it through the machine learning model.

The model is using a Support Vector Machine algorithm. The data is mapped to a high-dimensional feature space to categorize the data (IBM, 2021). The data is then transformed so that it can be plotted into one of two categories, edible or poisonous.

We use K-means clustering to visualize the two categories of edible and poisonous mushrooms and the attributes that cause a mushroom to be one or the other (descriptive method). We use a Support Vector Machine to predict the classification of mushroom (prescriptive method).

### Hypothesis verification

This product was designed with the assumption that a mushroom can be classified as edible or poisonous based on its attributes, and that these attribute combinations can be used to train a machine learning model to be able to identify mushrooms in this way. Using the initial dataset, we were able to confirm the hypothesis. Our model predicts the edibility of a mushroom with a 97% degree of confidence.

### Effective visualizations and reporting

This heatmap shows us the weight each attribute carries in respect to the other attributes. We can see from this image that the gill-size and bruises categories carry the most weight in determining a mushrooms class (edible or poisonous).

Chart, timeline

Description automatically generated

Our data is also closely distributed between poisonous and edible mushrooms.

Chart, pie chart

Description automatically generated

The confusion matrix for our SVM shows a low margin of error in predicting the mushrooms edibility.

Chart, treemap chart

Description automatically generated

Finally, we can plot the mushrooms on a scatter plot after performing Principal Component Analysis to condense the data into two-dimensional space.

Chart, scatter chart

Description automatically generated

### Accuracy analysis

The expected accuracy of the program was set to be 95% accurate. After training our model the program can predict the edibility of a mushroom with 97% accuracy. This exceeds our goal of 95% by 2%. We can consider this a high level of success. We could further improve accuracy by narrowing down the attributes we are training our model with to only those that hold a higher weight such as gill-size or bruises.

### Application testing

The program was tested during the implementation and testing phases of our project. Testing during the implementation phase allowed for us to fine-tune our model and user-interface before presenting it to the end-user.

During the testing phase we performed unit testing to verify the accuracy of our model. Then we performed user acceptance testing to ensure that the program would match the needs of the end-user/employees.

# Appendices

## Installation Guide

Prerequisites:

Anaconda 22.9.0

Python 3.7

Python Libraries:

* pandas 1.4.4
* numpy 1.21.5
* matplotlib 3.5.2
* seaborn 0.11.2
* pyinputplus 0.2.12
* ipywidgets 7.6.5
* scikit-learn 1.0.2
* jupyter 1.0.0

1. Install the prerequisites listed above
2. Extract the C964.zip folder into the directory you wish to run the program from
3. Open the Anaconda terminal and run the command “jupyter notebook”
4. In the jupyter notebook that opened in your web browser, find the directory you previously extracted C964.zip into
5. Open the Mushroom Classification.ipynb file
6. Click on “Kernel” > “Restart & Run All”

## User Guide

To start making predictions start the program and scroll to the bottom of the page until you see the column of dropdown menus.

A picture containing graphical user interface

Description automatically generated

Using each dropdown, select the matching value for each attribute of the mushroom you want to identify. Once you have finished entering the attributes click on the “Predict” button.

A picture containing text

Description automatically generated

This will display the predicted edibility that was determined with our model. To predict additional mushrooms all you need to do is change the attributes and click the “Predict” button again.

A picture containing graphical user interface

Description automatically generated

## Summation of Learning Experience

From this assignment I learned a multitude of things both related to machine learning and not. Learning to manipulate data and visualize it to tell a story is a skill that will be useful in my career moving forward. Using the various libraries like seaborn and matplotlib enabled me to visualize how our model would see the data and allowed me to visualize how each element relates to another.

I also learned about the value of planning out a project in advance. My first time doing this assignment I had an entirely different idea of what problem I wanted to solve, and the data I was going to use. I quickly found that it was a very intense undertaking for my first foray into machine learning. Instead, I eased into the process and found a friendlier dataset and another problem I was interested in solving.

If I were to restart this project, I would perform more data manipulation. Converting data to a more usable format sooner, as well as narrowing down the data elements that mean the most when training my model.

I learned a lot from this project, and from my time at Western Governors University as a whole. I am thankful for my time here, and for the assistance of the staff. I know that my progression and learning will continue past schooling and into my career.

## References

Gold, J. A. W., Kiernan , E., Yeh, M., Jackson, B. R., & Benedict, K. (2021, March 12). *Health care utilization and outcomes associated with accidental poisonous mushroom ingestions - United States, 2016-2018*. MMWR. Morbidity and mortality weekly report. Retrieved December 7, 2022, from <https://pubmed.ncbi.nlm.nih.gov/33705365/>

IBM. (2021, August 17). *How SVM Works*. How SVM works. Retrieved December 7, 2022, from https://www.ibm.com/docs/en/spss-modeler/saas?topic=models-how-svm-works

Synoptek. (2022, October 14). *Types of QA testing: Everything you need to know*. Synoptek. Retrieved December 7, 2022, from https://synoptek.com/insights/it-blogs/types-of-qa-testing-everything-you-need-to-know/

*Waterfall methodology*. IONOS Digital Guide. (2019, November 3). Retrieved December 7, 2022, from https://www.ionos.com/digitalguide/websites/web-development/waterfall-methodology/#:~:text=The%20five%2Dstage%20waterfall%20model,implementation%2C%20testing%2C%20and%20operation.