PROJECT PLAN

1. Students' name and Purdue e-mail.

a. Nikita Rajaneesh — nrajanee@purdue.edu

b. Swaraj Bhaduri — sbhadur@purdue.edu

c. Utkarsh Jain — jain192@purdue.edu

d. Ishaan Saxena — isaxena@purdue.edu

2. Definition of the problem, possibly relevant to your interests.

Analysis of Mushroom Species Data by classification into groups of poisonous and edible based on features such as caps, odor, stalks, etc.

3. Description of the dataset (or datasets) to be used.

Attribute Information: (classes: edible=e, poisonous=p)

1. **cap-shape**: bell=b, conical=c, convex=x, flat=f, knobbed=k, sunken=s
2. **cap-surface**: fibrous=f, grooves=g, scaly=y, smooth=s
3. **cap-color**: brown=n, buff=b, cinnamon=c, gray=g, green=r, pink=p, purple=u, red=e, white=w, yellow=y
4. **bruises**: bruises=t, no=f
5. **odor**: almond=a, anise=l, creosote=c, fishy=y, foul=f, musty=m, none=n, pungent=p, spicy=s
6. **gill-attachment**: attached=a, descending=d, free=f, notched=n
7. **gill-spacing**: close=c, crowded=w, distant=d
8. **gill-size**: broad=b, narrow=n
9. **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
10. **stalk-shape**: enlarging=e, tapering=t
11. **stalk-root**: bulbous=b, club=c, cup=u, equal=e, rhizomorphs=z, rooted=r, missing=?
12. **stalk-surface-above-ring**: fibrous=f, scaly=y, silky=k, smooth=s
13. **stalk-surface-below-ring**: fibrous=f, scaly=y, silky=k, smooth=s
14. **stalk-color-above-ring**: brown=n, buff=b, cinnamon=c, gray=g, orange=o, pink=p, red=e, white=w, yellow=y
15. **stalk-color-below-ring**: brown=n, buff=b, cinnamon=c, gray=g, orange=o, pink=p, red=e, white=w, yellow=y
16. **veil-type**: partial=p, universal=u
17. **veil-color**: brown=n, orange=o, white=w, yellow=y
18. **ring-number**: none=n, one=o, two=t
19. **ring-type**: cobwebby=c, evanescent=e, flaring=f, large=l, none=n, pendant=p, sheathing=s, zone=z
20. **spore-print-color**: black=k, brown=n, buff=b, chocolate=h, green=r, orange=o, purple=u, white=w, yellow=y
21. **population**: abundant=a, clustered=c, numerous=n, scattered=s, several=v, solitary=y
22. **habitat**: grasses=g, leaves=l, meadows=m, paths=p, urban=u, waste=w, woods=d

4. URL where the above dataset(s) is(are) available.

https://www.kaggle.com/uciml/mushroom-classification

5. Which machine learning algorithm(s) is(are) going to be used?

WRITE YOUR ANSWER HERE.

6. Cross-validation technique (e.g., training/validation/testing, k-fold cross-validation, bootstrapping)

WRITE YOUR ANSWER HERE.

7. Which hyperparameter(s) is(are) going to be tuned.

WRITE YOUR ANSWER HERE.

8. Description of the experimental results, e.g., plots of number of samples versus accuracy (you can use different subsets of the same dataset), regularization parameter versus accuracy, ROC curves, plots of different datasets, etc.

WRITE YOUR ANSWER HERE.

9. Which programming language are you going to use?

Python-2.7