Algorithm
Cat

Learning Today,
Leading Tomorrow

TASK LIST
Machine Learning

Crafting Careers, Shaping Futures

General Instructions

- All interns are requested to stay connected with our social Platforms (Telegram, LinkedIn & Instagram)
- The offer letter has officially landed in your inbox, accompanied by a power-packed task list tailored to each preferred domain.
- Upload your tasks on GitHub and post it on LinkedIn also tag @algorithmCat.
- To receive a "Recommendation letter" you must complete all tasks or else if you just want to receive an "Internship completion certificate" you can complete all tasks within the deadline.
- If you have any queries, contact us on email: algorithmcat.info@gmail.com

About Us

We are Algorithm Cat, a cool platform for students like you. We make internships easy and fun. You get firsthand experience, improve your skills, and boost your chances of landing a fantastic job. It is all about making your journey into the professional world impressive.

Roadmap of Internship

1. **LinkedIn Power Move: Optional, but why not? Post your offer letter on LinkedIn, tag "Algorithm Cat" for extra swagger.**

2. **Complete at least 6 tasks from the task list as per your choice.**

3. **Show and Tell: Share a video explanation or screenshot for each task on LinkedIn.**

4. **Submit the Task Submission form.**

5. **Epic Rewards: It is not just about the journey; it is about the destination. Receive a certificate after your stellar performance.**

Tasks

1. **TITANIC: It is your job to predict if a passenger survived the sinking of the Titanic or not. For each in the test set, you must predict a 0 or 1 value for the variable.**

2. **Your dataset had too many variables to wrap your head around, or even to print out nicely. How can you pare down this overwhelming amount of data to something you can understand?**

3. **The two datasets are related to red and white variants of the Portuguese "Vinho Verde" wine. For more details, consult the reference [Cortez et al., 2009]. Due to privacy and logistic issues, only physicochemical (inputs) and sensory (the output) variables are available (e.g., there is no data about grape types, wine brand, wine selling price, etc.).**

   These datasets can be viewed as classification or regression tasks. The classes are ordered and not balanced (e.g., there are much more normal wines than excellent or poor ones).

4. **The datasets consist of several medical predictor variables and one target variable, `Outcome`. Predictor variables includes the number of pregnancies the patient has had, their BMI, insulin level, age, and so on.**

5. **The SMS Spam Collection is a set of SMS tagged messages that have been collected for SMS Spam research. It contains one set of SMS messages in English of 5,574 messages, tagged according being ham (legitimate) or spam.**

MAJOR PROJECT: -

Project Description:

Sentiment analysis studies the subjective information in an expression, that is, the opinions, appraisals, emotions, or attitudes towards a topic, person, or entity. Expressions can be classified as positive, negative, or neutral. For example:

"I really like the new design of your website!" → Positive.

• You are provided the restaurant review datasets taken from Kaggle. Build a machine learning model (using NLP) to predict the label of the review either positive or negative. You are supposed to first process the data, then clean it using stemming/lemmatization, create a pipeline with Vectorization model and ML algorithm to predict the final sentiment.

• There are two features - 'review' - the sentence and 'sentiment' - the label for the review. 1 means positive review and 0 means negative review.

Get Data set from here

Algorithm

Cat

CHECK US OUT ON