BlockStak ML Engineer Assessment

Project Description:

Your task in this project is to collaborate with a team of data scientists to address a problem based on the provided dataset for BlockStak. The dataset details are provided at the end. We expect you to:

EDA:

- ☐ Study it closely and explore its details.
- ☐ Find out interesting information using visuals like charts and graphs.
- ☐ Explain what the data tells us using these visuals.

Predictive Modeling/Classification:

- Classify using a Decision Tree.
- Classify using Naive Bayes.
- Compare the results of the two techniques.

Conclusions and Recommendations:

- Detail your principal findings from the various sections.
- Provide actionable recommendations for the company to enhance their telemarketing strategy.

Optional (Not Mandatory): Develop an application utilizing your model with any Python framework, probably a web app using Django/DRF/Flask/FastAPI/StreamLit or others.

How to Submit Your Assessment:

1. **Review:** Double-check all requirements are met and code is commented.
2. **Problem Solving**: We're keen to understand your problem-solving approach. Highlight your thought process and strategies.
3. **Prepare Files:**
    - Use a public **GitHub** Repository **or** Combine files into a `.zip` **or** `.rar`**.**
    - Use a clear naming convention**, e.g., BlockstakML** (GitHub) **or** `YourName_BlockstakML.zip`**.**
4. **README:** Include a brief `README.md | README.txt` detailing your approach and any instructions for reviewers.
5. **Submit: Send us your file or link(s) in the email reply.**

Dataset Description:

This dataset pertains to the telemarketing endeavors of a bank. It's sourced from a Portuguese bank named “X”. The bank aims to refine its telemarketing strategy to promote long-term deposit accounts (like bonds and saving accounts). These marketing drives predominantly involved phone calls. Often, several interactions were necessary before determining if a customer would opt for a long-term deposit account. Your data science team will aid the bank by leveraging data analytics on the dataset provided.

1. **Age:** Customer's age (numeric).
2. **Job:** Job type (categorical).
3. **Marital:** Marital status (categorical).
4. **Education:** Customer's education level (categorical).
5. **Default:** Indicates if the customer has defaulted on credit (binary: "yes", "no").
6. **Balance:** Average yearly balance, in Euros (numeric).
7. **Housing:** Indicates if the customer has a housing loan (binary: "yes", "no").
8. **Loan:** Indicates if the customer has a personal loan (binary: "yes", "no").
9. **Contact:** Method of the last marketing campaign contact (categorical).
10. **Day:** Day of the month the customer was last contacted (numeric).
11. **Month:** Month of the year the customer was last contacted (categorical).
12. **Duration:** Duration of the last contact, in seconds (numeric).
13. **Campaign:** Number of contacts made during this marketing campaign for this customer (numeric).
14. **Pdays:** Days since the client's last contact from a prior campaign (numeric; -1 implies the client was never contacted before).
15. **Previous:** Number of contacts made before this campaign for this client (numeric).
16. **Poutcome:** Result of the previous marketing campaign (categorical).
17. **Y:** Class attribute denoting if the client subscribed to a term deposit (binary: "yes", "no").

Dataset:

https://docs.google.com/spreadsheets/d/1JmGVyfVgn2i0uhDZk4xc5kBDtlPqPnugT0HhcNYfNXY/edit?usp=sharing

Note: **Remember, exclude personal or sensitive data and follow all provided guidelines.**