

## MSDS 7346 Term Project Proposal

**Title:** "Utilizing AWS SageMaker for Advanced Data Analysis and Machine Learning Deployment"

### Team Members:

- Shikha Pandey, [pandey@smu.edu](mailto:pandey@smu.edu)
- Jessica Mcphaul, [jmcphaul@smu.edu](mailto:jmcphaul@smu.edu)

### I. Introduction and Connection to Data Science

This project aims to develop a comprehensive solution for real-time data analysis and predictive modeling using AWS SageMaker. It aligns with data science by leveraging cloud computing for efficient deployment and scalability of machine learning models.

### II. Specific Goals and Objectives

Our project has the following key goals:

1. **Efficient Setup and Use of SageMaker:** Establishing a robust SageMaker environment for machine learning workflows.
2. **Model Development and Deployment:** Developing and deploying a linear regression model for salary prediction.
3. **Comparative Analysis:** Evaluating SageMaker's performance against traditional machine learning approaches.

*Success Metrics:* Success will be measured by the smooth setup of SageMaker, accuracy of the predictive model, and efficiency gains over traditional methods.

### III. Detailed Methodology

Our approach includes:

1. **Tools & Technologies:** Utilizing AWS SageMaker, SageMaker Studio, and relevant AWS services.
2. **Data Handling:** Collecting and preprocessing salary data for EDA and model training.
3. **Step-by-Step Plan:** Sequentially executing project phases from initial setup to model deployment.

### IV. Feasibility and Project Scope

We anticipate challenges in configuring SageMaker and tuning the model for optimal performance. These will be addressed through extensive research and testing. The project is scoped to be achievable within two months.

### V. Literature Review

Preliminary literature review will focus on AWS SageMaker's capabilities and its applications in real-world scenarios. We aim to build upon existing knowledge while exploring new potentials in cloud-based machine learning.

### VI. Team Member Roles

- **Shikha Pandey:** Leading SageMaker setup and data preprocessing.
- **Jessica Mcphaul:** Focusing on model development and deployment.

### VII. Timeline and Milestones

- **Weeks 1-4:** Documentation Review and SageMaker Configuration
- **Week 5:** Data Collection and Preprocessing
- **Weeks 6-7:** Model Development and Testing
- **Week 8:** Model Deployment and Final Testing

### VIII. Expected Outcomes and Impact

We aim to successfully deploy a machine learning model using SageMaker, demonstrating its efficiency in cloud-based data science applications. This project is expected to contribute valuable insights into the use of cloud technologies in data science.

### IX. Ethical Considerations

We will ensure ethical handling of data, adhering to privacy standards and AWS usage policies.

### X. Contingency Plan

Should we encounter technical challenges with SageMaker, we plan to consult AWS documentation, seek expert advice, or consider alternative cloud platforms.

### XI. Conclusion

This project promises to enhance our understanding of AWS SageMaker and contribute to the field of data science, particularly in cloud-based machine learning applications. We are excited to explore and showcase the capabilities of SageMaker in our research.