Reflective Report

Process of Solving Problems and Learning to Use Jupyter Notebooks

At the beginning of this unit, my experience with Jupyter Notebooks was limited. However, through multiple portfolios, I significantly improved my skills, especially in data exploration, model application, and performance evaluation. By Portfolio 4, I had developed a more structured approach, efficiently handling tasks such as feature selection, model tuning, and cross-validation.

I faced challenges like managing different datasets, handling missing values, and selecting appropriate algorithms. Jupyter's flexibility allowed me to iteratively refine my models and document the process. By the end, I was confidently using advanced techniques like Recursive Feature Elimination (RFE) and cross-validation to optimize models.

Progress and Future Interests

Throughout the portfolios, I developed a better understanding of machine learning concepts. In Portfolio 1, I focused on basic tasks, while by Portfolio 4, I could handle more complex analyses, including distance metrics for a K-Nearest Neighbors (KNN) classifier and Logistic Regression. This progression reflects my growing ability to derive insights from data.

In the future, I aim to apply these skills to real-world problems, particularly in healthcare or environmental sustainability, where predictive models can help with early intervention and decision-making.

Discussion Points Based on Portfolio 4

1. **Why I Chose the Dataset**

For Portfolio 4, I chose the **"Estimation of Obesity Levels Based on Eating Habits and Physical Condition"** dataset. Obesity is a global health concern, and the dataset offered diverse features like demographics, eating habits, and physical activity. These factors provided a rich foundation for exploring classification models to predict obesity risk.

2. **Identifying the Problem**

The problem I aimed to solve was predicting obesity based on lifestyle factors. Using machine learning, I sought to uncover patterns in the data that could indicate high-risk individuals, relevant to public health interventions.

3. **Choosing Machine Learning Models**

I applied two models: **K-Nearest Neighbors (KNN)** and **Logistic Regression**. KNN, with its focus on proximity, was ideal for this dataset, achieving 92% accuracy. Logistic Regression provided a probabilistic framework, helping identify key features, such as snack frequency and family history of obesity.

4. **Insights and Conclusion**

The models confirmed that factors like age, vegetable consumption, and physical activity significantly impact obesity risk. Additionally, small lifestyle habits, such as snack frequency and transportation methods, had an unexpected influence, emphasizing the importance of practical interventions.

KNN outperformed Logistic Regression in terms of accuracy, but both models offered valuable insights into the relationships between lifestyle factors and obesity.

5. **Decision Tree Analysis**

I also explored **Decision Tree Analysis** in Portfolio 4. Decision trees helped visualize the most influential factors in predicting obesity, like snack frequency and vegetable consumption. However, decision trees tend to overfit, so I used **pruning techniques** to simplify the model and enhance generalization. This experience has motivated me to explore **ensemble methods** like **Random Forests**, which offer improved performance by combining decision trees.

Conclusion

My journey through the portfolios has provided a solid foundation in data science and machine learning. I've learned how to preprocess data, apply different algorithms, and extract meaningful insights. Moving forward, I'm excited to apply these skills to real-world challenges, particularly in healthcare, where data-driven approaches can make ats. Let me know if you need further adjustments!ough pruning. Let me know if you need further adjustments!significant impact on improving outcomes. a tangible impact on improving outcomes.

In [ ]:

```

```