Data Mining

Assignment 1- A case of model overfitting

Scenario:

A retail company wants to predict sales trends to boost profits and manage inventory better. They decide to use a complex neural network model to forecast sales of individual products across various outlets. This model considers factors like seasonal patterns, promotions, demographics, and even weather conditions, trained on past sales data.

Factors Causing Overfitting:

- **Model Complexity:** The neural network has many layers and features, making it good at capturing complex patterns, including noise and outliers in the training data, leading to overfitting.
- **Limited Data:** Despite trying to gather lots of data, the dataset might still be small, making it hard for the model to adapt to new data.
- **Feature Engineering:** Adding too many features designed for the training data or irrelevant ones can also cause overfitting.

Prevention of Overfitting:

- **Keep It Simple:** Instead of complex neural networks, simpler models like decision trees or linear regression can work better when data is limited.
- **Regularization Techniques:** Techniques like dropout, early stopping, or regularization can help limit the model's capacity to fit noise in the data.

- **Cross-Validation:** Splitting data into groups for training and validation helps check the model's performance on new data and spot overfitting.

Actions to Avoid Overfitting:

- **Choose Features Wisely:** Select features carefully based on knowledge of the domain and smart data analysis to prevent the model from fitting noise.

- **Data Expansion:** Create new datasets or add to existing ones to give the model more samples to learn from and reduce overfitting.

- **Ensemble Methods:** Combining predictions from multiple models often reduces overfitting and improves generalization compared to using just one model.

References:

Lewis, M. R. (2024, March 19). How to Write a Use Case: 10 Steps (with Pictures) - wikiHow. wikiHow. https://www.wikihow.com/Write-a-Use-Case

EliteDataScience. (2022, July 6). Overfitting in Machine learning: What it is and how to prevent it. EliteDataScience. https://elitedatascience.com/overfitting-in-machine-learning