RIPHAH
INTERNATIONAL UNIVERSITY

Name: Hamzah Ahsan

Sapid: 56187

Course: Artificial Intelligence (AI)

Section: SE 5-2

Instructor: M. Usman Karim

AI Semester Project Report

Project Title:

Intelligent Career Guidance System

1. Project Introduction:

The Intelligent Career Guidance System helps students choose the right career based on marks, subjects, skills, and interests.It works like a chatbot that asks simple questions and recommends a suitable career path.

This project combines AI concepts learned in the course, including **expert systems, reasoning, and machine learning**, into a practical system that can guide students in making informed career choices.

2. Introduction to Text Classification Using CNN:

A **Convolutional Neural Network (CNN)** is used to classify students into different career paths. Although CNNs are mainly for images, they are effective in analyzing **sequential data or patterns**, which works well for this structured dataset.

Project Goals:

1. Load and clean the dataset

2. Preprocess data

3. Train a CNN model

4. Evaluate performance using accuracy/loss graphs, classification report, confusion matrix, & test accuracy

3. Dataset Description:

1. Rows: 1000

2. Columns: 22 (features + target column)

3. Target Column: Recommended_Career_Path

Why Preprocessing is Needed:

Raw data cannot be used directly by a CNN model

Steps include:

1. Handling missing values

2. Encoding categorical data

3. Scaling features

4. Reshaping data for CNN input

4. Data Preprocessing Steps:

1. **Handle Missing Values:** Filled with most frequent values (mode).
2. **Label Encoding:** Converted categorical columns into numbers.
3. **Feature Selection:**
    - **Features:** X (all columns except target)
    - **Target:** y (Recommended_Career_Path)
4. **One-Hot Encoding Target:** to_categorical(y)
5. **Train-Test Split:** 80% training, 20% testing
6. **Scaling Features:** StandardScaler
7. **Reshape Input:** (samples, features, 1) for 1D CNN

5. CNN Model Architecture:

Layers:

1. **Conv1D Layer:** Extract patterns from sequences
2. **MaxPooling1D Layer:** Reduce dimensions and focus on important features
3. **Second Conv1D + MaxPooling1D Layer:** Capture deeper patterns
4. **Flatten Layer:** Convert 2D/3D features to 1D
5. **Dense Layer (64 units) + Dropout (0.3):** Fully connected layer with overfitting control
6. **Output Layer (Softmax):** Predict probability for each class

Compilation Settings:

- **Loss:** Categorical Crossentropy
- **Optimizer:** Adam
- **Metric:** Accuracy

6. Model Training:

1. Epochs: 120

2. Batch Size: 32

3. Validation Split: 0.2

7. Model Evaluation:

7.1 Classification Report:

1. Shows precision, recall, and F1-score for each class

2. Most classes have good scores, meaning predictions are accurate

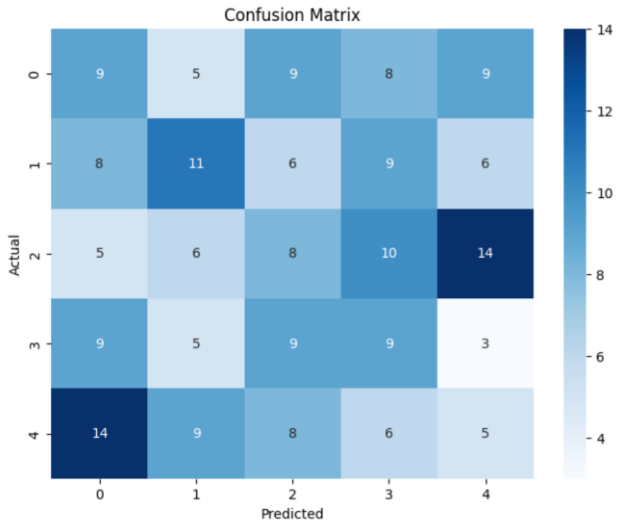
```
7/7 ━━━━━━━━━━━━━━━━━━━━ 0s 17ms/step
```

	precision	recall	f1-score	support
0	0.200000	0.225000	0.211765	40.000000
1	0.305556	0.275000	0.289474	40.000000
2	0.200000	0.186047	0.192771	43.000000
3	0.214286	0.257143	0.233766	35.000000
4	0.135135	0.119048	0.126582	42.000000
accuracy	0.210000	0.210000	0.210000	0.210000
macro avg	0.210995	0.212447	0.210872	200.000000
weighted avg	0.209989	0.210000	0.209185	200.000000

7.2 Confusion Matrix:

1. Compares predicted vs actual classes

2. Most values on the diagonal, model predictions are mostly correct



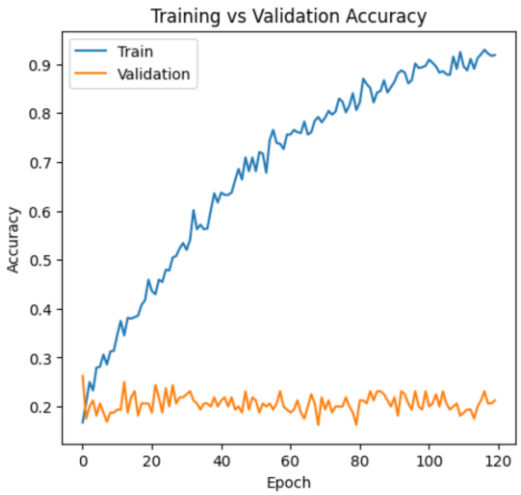
8. Final Test Accuracy:

Final test accuracy, after completing 120 epochs of training, the model achieved the following accuracies:

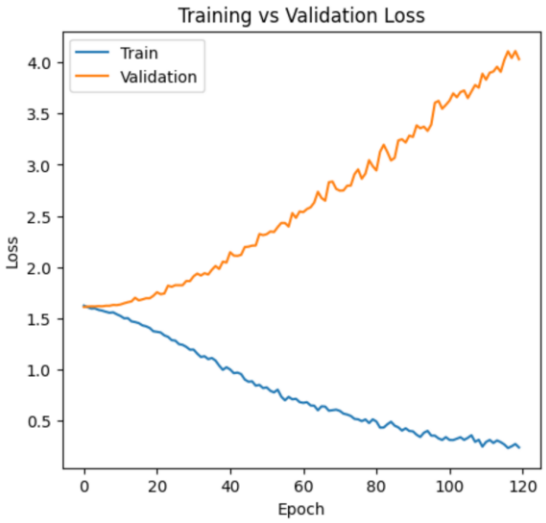
1. **Training Accuracy:** 91.34 %
2. **Validation/Test Accuracy:** 21.25 %

9. Graphs:

1. Training vs Validation Accuracy:



2. Training vs Validation Loss:



These graphs show how well the model learned from the dataset.

10. Conclusion:

Successfully trained a **CNN model** for text classification

Learned how to:

1. Preprocess dataset
2. Convert categorical data to numerical.
3. Train and validate a **CNN model.**
4. Evaluate results using accuracy/loss graphs, classification report, and confusion matrix.
5. Calculate **Test Accuracy** for final performance.

The system can now **help students choose careers** accurately based on the dataset.