API Documentation for Traffic Prediction Model

Overview

The Traffic Prediction Model API is designed to predict traffic conditions using machine learning models (LSTM/GRU). The API is built using Flask and deployed in a Kubernetes environment. The primary function of this API is to receive time-series data as input and return predictions that indicate the traffic conditions based on historical data.

Endpoint Descriptions

1. /predict Endpoint

- **HTTP Method**: POST
- **Description**: This endpoint accepts a JSON payload containing a batch of time-series data and returns predictions generated by the machine learning model.
- **Content-Type**: application/json

Request Body

- The request body should contain a JSON object with the following structure:

```
{
    "input": [
        [
            [value1, value2, value3],
            [value4, value5, value6],
            ...
        ],
        [
            [value7, value8, value9],
            [value10, value11, value12],
            ...
        ]
```

```
    ]
}
```

- **input**: A list of time-series sequences where each sequence is a list of numerical feature vectors representing different time steps. The shape should be (batch_size, time_steps, num_features).

Response Format

- The response is a JSON object containing the predictions for each input sequence:

```
{
    "prediction": [
        [prediction_value1],
        [prediction_value2],
        ...
    ]
}
```

- **prediction**: A list of prediction values corresponding to each input sequence.

Sample Requests and Responses

Sample Request

Here is an example of a POST request to the /predict endpoint:

HTTP Method: POST

Headers: Content-Type: application/json

Request URL: http://<your -ip>:5000/predict

Request Body:

```
{
    "input": [
        [
            [0.123, 0.234, 0.345],
            [0.456, 0.567, 0.678],
            [0.789, 0.890, 0.901]
        ],
        [
            [0.223, 0.334, 0.445],
            [0.556, 0.667, 0.778],
            [0.889, 0.990, 1.101]
        ]
    ]
}
```

Sample Response

The response from the server will look like the following:


```
{
    "prediction": [
        0.3357592523097992,
        0.9687418937683105,
        0.9758992195129395,
        0.29004836082458496,
        0.9763253927230835,
        0.3591136336326599,
        0.9374858140945435,
        0.8968163728713989,
        1.0043978691101074,
        0.9506070613861084,
        0.97392338514328,
        0.9382553100585938
    ]
}
```


This indicates that the model has processed the input data and provided predictions for each sequence in the request.

Error Handling

Common Error Messages

1. **Invalid Input Format**
    - **Error Response**:


```
{
    "error": "Invalid input format. Please ensure the input data is a 3D list with shape (samples, time_steps, features)."
}
```

- **Cause**: This error occurs when the input data does not match the expected shape or structure.
- **Solution**: Verify that the input data is properly structured as a batch of time-series sequences.

2. **Model Prediction Error**
   - **Error Response**:

```
{
    "error": "An error occurred during prediction."
}
```

   - **Cause**: This error may happen if there is a problem with the model's prediction process or if the input data shape does not match the model's requirements.
   - **Solution**: Ensure that the input data shape matches the training data's format and that the data types are correct.

3. **Internal Server Error**
   - **Error Response**:

```
{
  "error": "Internal server error. Please try again later."
}
```

   - **Cause**: This error indicates a server-side issue that prevents the application from responding correctly.
   - **Solution**: If this error persists, contact the API support team or check the server logs for more details.

Conclusion

This API provides a reliable way to interact with the traffic prediction model for forecasting traffic conditions. By using the /predict endpoint, you can send batches of time-series data and receive predictions in real-time, enabling integration with traffic management systems or analytics platforms.

For questions or further assistance regarding the API, please contact the support team.