**Job Specification: NLP and Maritime Geospatial Data Scientist**

**Position Title:**

**NLP & Maritime Geospatial Data Scientist / Engineer**

**Department:**

Maritime Domain Awareness (MDA) / Data Analytics Division

**Reporting To:**

AVP (Strategic Projects) / Head – AI & Analytics

**1. Position Overview**

The position involves designing and implementing an **NLP-driven analytical interface** to interact with a **maritime geospatial database** containing **AIS transmissions** and related datasets.  
The incumbent will be responsible for developing models and systems capable of:

* Understanding natural language queries related to vessel movements,
* Translating them into structured geospatial analytics, and
* Providing predictive insights and anomaly detection on vessel trajectories.

This role requires a multidisciplinary skill set combining **Natural Language Processing (NLP)**, **Geospatial Data Science**, and **Predictive Modeling** within the **maritime domain**.

**2. Key Responsibilities**

**A. NLP System Design & Development**

* Develop a **natural language interface** that enables users to query vessel positions, predict future locations, and verify trajectory consistency using plain language inputs.
* Design and train **intent recognition and entity extraction** models specific to maritime terminology (e.g., MMSI, SOG, COG, EEZ, route deviation).
* Implement **semantic mapping** from user queries to structured database or analytic operations.

**B. Data Integration & Management**

* Work with **AIS data streams** (live and historical) and integrate with a **PostGIS-based geospatial database**.
* Design and maintain efficient data pipelines for **ETL, cleaning, normalization, and time-series storage** of AIS data.
* Implement **spatial-temporal indexing and optimization** for rapid retrieval and analysis.

**C. Predictive Analytics & Anomaly Detection**

* Develop and validate **vessel trajectory prediction models** using statistical and machine learning methods (Kalman Filters, ARIMA, LSTM, etc.).
* Implement **trajectory consistency checks** to verify whether a vessel’s latest reported position aligns with its historical movement pattern.
* Create **anomaly detection algorithms** for identifying sudden deviations, spoofing, or inconsistent data patterns.

**D. API & Integration**

* Develop backend **APIs** to expose NLP query results and predictive analytics to client applications or dashboards.
* Work closely with **frontend developers** to enable geospatial visualization and interactive reporting.

**E. Research & Continuous Improvement**

* Stay updated with current developments in **maritime analytics**, **LLM-based query systems**, and **geospatial AI**.
* Explore integration of contextual data (weather, currents, port congestion) to enhance predictive accuracy.
* Prepare documentation, white papers, and demonstration modules for internal and client review.

**. Required Technical Skills**

| **Category** | **Required Skills / Tools** |
| --- | --- |
| **Programming & Data** | Python (Pandas, NumPy, GeoPandas, scikit-learn), SQL |
| **NLP Tools** | spaCy, Hugging Face Transformers, NLTK, LangChain or OpenAI API |
| **Machine Learning / Prediction** | Kalman Filters, LSTM/GRU, ARIMA, anomaly detection methods |
| **Geospatial Technologies** | PostgreSQL + PostGIS, Shapely, Folium, QGIS, CesiumJS or Leaflet |
| **Visualization** | Matplotlib, Plotly, or equivalent geospatial visualization frameworks |
| **APIs & Backend** | FastAPI / Flask for query-response systems |
| **Version Control** | Git / GitHub |
| **Preferred (Domain)** | Knowledge of AIS data structure, vessel behavior, maritime zones, or MDA systems |

**4. Educational Qualifications**

* **Master’s degree** in Computer Science, Data Science, Geoinformatics, Ocean Engineering, or a related field.
* Candidates with **Bachelor’s degree + significant relevant experience** will also be considered.

**5. Experience Requirements**

* **3–6 years** of hands-on experience in:
  + NLP systems or conversational AI projects.
  + Geospatial or temporal data analytics.
  + Predictive modeling using time-series data.
* Prior experience in **maritime analytics, defense systems, or MDA platforms** is highly desirable.

**6. Key Competencies**

* Strong analytical and problem-solving mindset.
* Ability to design modular, scalable systems combining AI and GIS.
* Familiarity with maritime operational context (AIS, ports, EEZ, vessel tracking).
* Excellent documentation, communication, and presentation skills.
* Self-driven, with capacity to work independently and collaboratively across technical teams.

**7. Performance Metrics**

* Accuracy and robustness of NLP query translation.
* Prediction accuracy (e.g., mean positional error in vessel forecast).
* Response time and scalability of the database-querying layer.
* Quality of anomaly detection and validation against real AIS tracks.
* Integration success within broader MDA software ecosystems.

**8. Employment Type & Location**

* **Full-time**, based at the company’s **Delhi development centre**.
* May require occasional interaction with **Navy, Coast Guard, or maritime authorities** for data understanding and validation.

**Programming Task: AIS Data NLP Query Prototype**

**Objective**

Build a **small working prototype** that accepts a **natural language query** about vessel movement and returns an **analytical or predictive output** using a provided AIS dataset.

**Dataset –** Open SourceDataset link

<https://datasetsearch.research.google.com/search?query=ais%20vessel%20data>

<https://coast.noaa.gov/htdata/CMSP/AISDataHandler/2020/index.html>

**Task Requirements**

**1. NLP Interpretation**

* Accept a **natural language query** such as:
  + “Show the last known position of INS Kolkata.”
  + “Predict where MSC Flaminia will be after 30 minutes.”
  + “Check if the latest position of Ever Given is consistent with its past movement.”
* Parse the query to extract:
  + **Intent** (show, predict, verify)
  + **Vessel name**
  + **Time horizon** (if applicable)

*(Can use simple keyword mapping or a lightweight NLP model like spaCy; no need for heavy LLM integration.)*

**2. Data Query & Processing**

* Read the AIS CSV into a dataframe.
* Based on extracted intent:
  + **If “show”** → return the latest known lat/lon for that vessel.
  + **If “predict”** → estimate the vessel’s next position after X minutes assuming constant speed and course.
  + **If “ Verify : plot the course last 5 points and estimated 30 minutes interval (5 minutes In timestamp -)**

**3. Output**

* Print or return a short textual response like:
* INS Kolkata was last seen at 13.01N, 80.27E moving SE at 15 knots.
* Predicted position after 30 minutes: 12.92N, 80.55E.
* Optional bonus: plot vessel track on a simple map (using Folium or Matplotlib).

**4. Evaluation Criteria**

| **Skill Area** | **What You’re Looking For** |
| --- | --- |
| **Data Handling** | Clean reading and processing of AIS data. |
| **Logic & Algorithms** | Correct computation of vessel heading, position prediction, and movement validation. |
| **NLP Understanding** | Ability to extract intent and entities from text (even with simple logic). |
| **Code Structure** | Modular, readable, and documented code. |
| **Domain Sense** | Awareness of realistic vessel speeds and behaviors. |

**Expected Deliverables**

1. Python script or notebook (ais\_nlp\_query.py or .ipynb)
2. Sample output for 2–3 example queries.
3. Short README explaining:
   * How to run the code
   * NLP approach used
   * Prediction logic