Project 2: Cloud Document Database with Predictive Analytics

MongoDB/Cosmos DB on Azure + Neural Network Modeling + Interactive Dashboard

This is an individual project with a due date of Saturday, November 22nd (7:30 AM). The presentation is the same day from 8:00 AM to 10:00 AM.

In this comprehensive project, you will combine cloud-based document database management with advanced predictive modeling and interactive visualization. You will deploy a MongoDB or Cosmos DB database on Microsoft Azure, develop and evaluate predictive models (including at least one neural network), create an interactive dashboard for data exploration, and deliver a professional 5-minute pitch presenting your solution.

Part 1: Cloud Document Database

Deploy a document-based database on Microsoft Azure using MongoDB or Azure Cosmos DB. Your database must be accessible online and properly configured for secure remote connections.

You may use existing document databases or create your own from structured data. Ensure the data is substantial enough to support meaningful predictive modeling. Suggested domains:

- E-commerce (product catalogs, customer transactions, reviews)
- Healthcare (patient records, treatment outcomes, medical imaging metadata)
- IoT sensor data (environmental monitoring, smart devices, telemetry)
- Social media analytics (posts, engagement metrics, user profiles)
- Education (student performance, course enrollment, learning analytics)

Part 2: Predictive Modeling

Develop and evaluate multiple predictive models using data from your document database. At least one model <u>must be a neural network</u>. Your analysis should demonstrate systematic model comparison and justify your final model selection.

- Develop at least three different predictive models
- At least one model must be a neural network (feedforward, CNN, RNN, LSTM, or transformer-based)
- Include at least one traditional machine learning model for comparison (Random Forest, Gradient Boosting, Logistic Regression, SVM)
- Properly split data into training, validation, and test sets
- Implement appropriate evaluation metrics for your problem type (classification, regression, time series, etc.)
- Document hyperparameter tuning process and final model selection rationale

Your neural network implementation should include:

- Clear architecture description (layers, neurons, activation functions)
- Appropriate loss function and optimization algorithm
- Training history visualization (loss curves, accuracy curves)
- Techniques to prevent overfitting (dropout, regularization, early stopping)
- You may use TensorFlow, Keras, PyTorch, or similar frameworks

Model Comparison and Analysis

Provide a comprehensive comparison of all models including:

- Performance metrics table comparing all models
- Analysis of training time and computational requirements
- Discussion of interpretability vs. performance tradeoffs
- Final recommendation with justification based on business context

Part 3: Interactive Dashboard

Design and build an interactive dashboard that connects to your Azure database and enables data exploration, model insights, and decision support. The dashboard can run locally, but must connect to your online database.

Requirements:

- Connect directly to your Azure MongoDB/Cosmos DB using secure connection strings
- Include at least 4-5 interactive visualizations showing key metrics and trends
- Incorporate model predictions or insights into the dashboard
- Implement user controls (filters, dropdowns, sliders) for data exploration
- Use Streamlit, Dash, Shiny, or a similar framework
- Ensure professional design with clear labels, appropriate color schemes, and logical layout

Part 4: Written Narrative

Prepare a comprehensive written document (3-5 pages) that explains your project in a professional context. This narrative should be accessible to both technical and non-technical stakeholders.

Your narrative must include:

- Problem Statement: Clear definition of the business or social problem being addressed
- Data Context: Description of data sources, database structure, and relevance to the problem
- Modeling Approach: Explanation of models developed, why they were chosen, and how they were evaluated
- Dashboard Purpose: How the dashboard supports decision-making and who the intended users are
- Business Value: Discussion of how this solution creates value in realistic business or social context (retail, healthcare, education, IoT, etc.)
- Technical Implementation: Brief overview of Azure configuration, modeling frameworks used, and dashboard technology

Part 5: Project Pitch

Deliver a professional 5-minute presentation pitching your solution to a hypothetical stakeholder group (executives, investors, or clients). This presentation should focus on business value and practical impact rather than technical minutiae.

Pitch Structure Guidelines:

- Opening (30-45 seconds): Hook your audience with the problem and its impact
- Problem Context (1 minute): Clearly articulate the business challenge or opportunity
- Solution Overview (1.5 minutes): Explain your approach at a high level: database architecture, predictive models, and dashboard
- Live Demo (1.5 minutes): Show your dashboard in action, highlighting 2-3 key insights or predictions
- Value Proposition & Close (30-45 seconds): Emphasize ROI, risk reduction, efficiency gains, or other business benefits

Presentation Tips:

- Use slides with minimal text and strong visuals
- Practice your timing to stay within 5 minutes
- Focus on business outcomes rather than technical implementation details