Statement of Work (SOW)

Project Title: Personalized Health Insurance Premium Estimator
Document Version: 1.0
Date: June 10, 2025
SOW Reference: SOW-SI-2025-001

---

1. Project Overview

Client: Shield Insurance
Service Provider: AtliQ AI
Project Manager: Bharathkumar Tamilarasu
Data Scientist: Bharathkumar Tamilarasu

AtliQ AI will develop and deploy a machine learning-based predictive model for Shield Insurance to accurately estimate health insurance premiums based on demographic, lifestyle, and medical history factors. The solution will include a cloud-deployed model with an interactive web application for insurance underwriters.

---

2. Business Objectives

Primary Objectives:

- **Accuracy Target:** Achieve minimum 97% prediction accuracy on validation dataset
- **Precision Target:** Ensure 95% of incorrect predictions deviate by ≤10% from actual values
- **User Adoption:** Enable 100% of underwriters to use the system within 30 days of deployment

Success Metrics:

- Model performance metrics (R^2, MAE, RMSE)
- Processing time <3 seconds per prediction

---

3. Scope of Work

3.1 Data Cleaning & Analysis

Duration: 2 weeks

- **Data Collection:** Acquire and validate datasets (minimum 10,000 records)
    - Demographics: Age, gender, location, income, marital status
    - Health metrics: BMI, medical history, family history
    - Lifestyle factors: Smoking status
- **Data Quality Assessment:** Completeness, accuracy, consistency checks
- **Data Preprocessing:** Cleaning, normalization, handling missing values, outlier detection
- **Exploratory Data Analysis:** Statistical analysis, correlation studies

3.2 Model Development & Optimization

Duration: 2 weeks

- **Feature Engineering:** Create derived features, encode text features
- **Algorithm Selection:** Evaluate multiple ML algorithms:
    - Linear/Polynomial Regression
    - Random Forest
    - Gradient Boosting (XGBoost)
- **Model Training:** Cross-validation, hyperparameter tuning using GridSearch/RandomSearch
- **Performance Evaluation:** Comprehensive testing on train/validation/test splits
- **Error Analysis:** Analyse prediction errors to ensure 95% of incorrect predictions deviate by ≤10% from actual values, identify error patterns across customer segments
- **Model Selection:** Choose optimal model based on accuracy, interpretability, and business requirements

3.3 Streamlit Application Development

Duration: 1 week

- **User Interface Design:** Intuitive form-based input interface
- **Core Features:**
    - Real-time premium prediction
    - Batch processing capability

3.5 Testing & Quality Assurance

Duration: 1 week

- **Integration Testing:** End-to-end workflow validation
- **Regression Testing:** Ensure model stability across different data scenarios

3.6 Documentation

Duration: 1 week

- **Technical Documentation:** Architecture, deployment guides
- **User Manual:** Step-by-step guides with screenshots and examples

---

4. Deliverables

4.1 Technical Deliverables

- **Trained ML Model:** Serialized model files
- **Streamlit Application:** Fully functional web application
- **Source Code:** Complete codebase with version control (Git repository)

4.2 Documentation Deliverables

- **Technical Architecture Document:** System design and component interactions
- **User Manual:** End-user guide with screenshots and workflows
- **Deployment Guide:** Step-by-step infrastructure setup instructions

---

5. Project Timeline

Phase	Duration	Start Date	End Date	Key Milestones
Phase 1: Data Cleaning & Analysis	2 weeks	Week 1	Week 2	Data validation complete
Phase 2: Model Development & Optimization	2 weeks	Week 3	Week 4	Model accuracy target achieved
Phase 3: Application Development	2 weeks	Week 5	Week 6	Streamlit app functional
Phase 4: Testing	1 week	Week 7	Week 7	All tests passed
Phase 5: Documentation	1 week	Week 8	Week 8	Well documented details

Total Project Duration: 8 weeks

---

6. Risk Management

6.1 Technical Risks

Risk	Probability	Impact	Mitigation Strategy
Model accuracy below 97%	Medium	High	Extended model tuning phase, feature engineering
Cloud deployment issues	Low	Medium	Thorough testing in staging environment
Data quality problems	Medium	High	Early data validation

---

7. Communication Plan

7.1 Reporting Structure

- **Weekly Status Reports:** Every week on Friday

---

8. Acceptance Criteria

8.1 Technical Acceptance

- Model achieves ≥ 97% accuracy on validation dataset
- 95% of predictions within 10% of actual values
- Model Prediction time <3 seconds

8.2 Business Acceptance

- Underwriters can successfully use the system independently
- All required documentation provided and approved

---

9. Signatures

Shield Insurance
Representative: _____
Title: _____
Date: _____
Signature: _____

AtliQ AI
Representative: Bharathkumar Tamilarasu
Title: Project Manager & Data Scientist
Date: _____
Signature: _____

Document Control:
Created: June 10, 2025
Last Modified: June 10, 2025
Version: 1.0