**DS Project Scoping Document: STELLANTIS 2025**

**General guidelines / Notes and tips for writing a project scoping document**

A few pieces of advice to create a document that will better serve the goals of the “product scoping document”:

* Keep the document short, in bullets, and to the point.
* Keep it clean, clear and readable. Something a stakeholder or a colleague can read fast and understand the main points
* The document is **dynamic** and might change over time. That's ok. Just inform relevant parties on big changes
* Share this document with all clients and stakeholders, every party that you have a requirement from and with your teammates, let them know what you are doing.

**Template**

1. **Intro section:**
   1. High level project description (TL;DR - goal, client, companies involved..)
   2. Timeline of major milestones: algo review, presentation to stakeholders, deployment, A/B test
   3. Team (data scientists, engineers, labelers, product owner…)

1. **Product section**
   1. Who is the point of contact?
   2. Who is the client?
   3. What is the client’s pain?
   4. What the model should give / how should it solve the problem?
   5. Where will the model be deployed?
   6. What are the business KPIs
   7. What is the expected/desired business impact?

1. **Data section**
   1. Time period
   2. Amount of data
   3. Population
   4. Labels
   5. Train / test sets

1. **Solution Section**
   1. Baseline model / version / heuristic to compare with
   2. Initial modeling approach
   3. Solution assumptions and risks
      1. What are the product / data assumptions?
      2. What is the risk from these assumptions?
      3. Risks mapping (research / model / data, engineering, business)
   4. Evaluation metrics

1. **Requirements section**
   1. Dataops
      1. labeling & data curation efforts
      2. Datadev needed to get / process the data
   2. Engineering requirements
      1. Backend / frontend dev need to integrate or deploy the model
      2. Model consumption and API
   3. A/B test - notes / requirements

1. **Legal consult requirements**
   1. Data usage that need legal approval (in terms of GDPR, controller-processor relation, PII, 3rd party data, open datasets)
   2. Open source / packages in use (verify licencing etc)
   3. AI regulations

**STELLANTIS 2025: Enhancing ETA and Energy Consumption Predictions through Driving Pattern Analysis**

**Intro Section**

* **Project Description:** Enhance Stellantis' ETA and energy consumption prediction models by integrating driving behavior patterns derived from vehicle sensor data.
* **Timeline:**
  + Algorithm review
  + Presentation to stakeholders
  + Deployment
  + A/B testing
* **Team:** Data scientists, engineers, labelers, product owner.

**Product Section**

* **Client:** Stellantis
* **Client’s Pain:** Current ETA and energy consumption predictions do not incorporate driving behavior, potentially reducing accuracy.
* **Model Output:** A classifier to identify driving patterns and features that improve ETA and energy consumption predictions.
* **Deployment:** Stellantis' vehicle navigation or telematics systems.
* **Business KPIs:** Increased accuracy of ETA and energy consumption predictions.
* **Business Impact:** More personalized and efficient navigation solutions, reinforcing Stellantis’ focus on innovation and sustainability.

**Data Section**

* **Data:** Time-series vehicle sensor data (coordinates, location, temperature, altitude); initial use of a public dataset (e.g., Kaggle), followed by proprietary data.
* **Population:** Subset of vehicles or a specific geographic region.
* **Labels:** Driving pattern labels to be defined and curated.
* **Train/Test Sets:** To be generated from the dataset.

**Solution Section**

* **Baseline:** Existing ETA and energy consumption models without driving pattern features.
* **Approach:** Create a driving style feature space, develop a classifier for pattern identification, validate on public data, then apply to proprietary data to enhance predictions.
* **Assumptions:**
  + Driving patterns influence ETA and energy consumption.
  + Sensor data adequately captures these patterns.
* **Risks:**
  + Data quality issues (e.g., missing values).
  + Challenges in defining and labeling driving patterns.
  + Computational complexity.
* **Evaluation Metrics:**
  + Classifier: Accuracy, F1-score.
  + Prediction improvement: MAE, RMSE.
  + Correlation between patterns and trip metrics.

**Requirements Section**

* **Dataops:** Clean data, address missing values, perform feature engineering.
* **Labeling:** Curate and label driving patterns.
* **Datadev:** Process public and proprietary datasets.
* **Engineering:** Integrate model into production systems.
* **Dev Needs:** Backend and frontend development for deployment.
* **Model API:** Define how the model will be accessed.
* **A/B Test:** Assess prediction improvements in production.
* **Legal:**
  + Ensure GDPR compliance for sensor data (e.g., location as PII).
  + Verify licenses for open-source tools or datasets.

**1. Intro Section**

* **High-Level Description**
  + **Project Title**: Enhancing ETA and Energy Consumption Predictions Through Driving Pattern Analysis
  + **Goal**: Incorporate driving behavior patterns into current ETA and energy consumption models, aiming for improved prediction accuracy and user experience.
  + **Client / Companies Involved**: Stellantis (automotive manufacturer); potential synergy with external data providers.
* **Timeline (Tentative)**
  + **Phase 1**: Data collection/cleaning & exploration (Weeks 1–4)
  + **Phase 2**: Feature engineering (Weeks 5–8)
  + **Phase 3**: Classification model development & validation (Weeks 9–12)
  + **Phase 4**: Integration into ETA & consumption prediction (Weeks 13–16)
  + **Phase 5**: Final review, presentation to stakeholders, potential A/B testing (Weeks 17–20)
* **Team**
  + **Data Scientists**: Responsible for feature engineering, model development, evaluation
  + **Data Engineers**: Handle data ingestion, processing pipelines, infrastructure
  + **Product Owner / Stakeholder**: Provides business requirements, signs off on deliverables
  + **Engineers / DevOps**: Integrate model into production, support deployment

**2. Product Section**

* **Point of Contact**: [Name/Title at Stellantis or internal champion]
* **Client**: Stellantis (internal stakeholders in R&D or navigation services)
* **Client’s Pain**:
  + Current ETA and energy consumption models lack driving-behavior-based features.
  + Inaccurate predictions affect user satisfaction and route planning efficiency.
* **Proposed Solution**:
  + Identify and model driving style patterns (e.g., acceleration, braking, average speed fluctuations) using vehicle sensor data.
  + Integrate these patterns as features in ETA and energy consumption models.
* **Deployment**:
  + On Stellantis production servers or cloud-based environment.
  + Consumed by in-vehicle systems (e.g., infotainment, navigation) or telematics platforms.
* **Business KPIs**:
  + Prediction accuracy (reduced ETA error, more precise energy consumption estimates).
  + Improved user satisfaction/retention.
  + Potential cost savings (fuel/battery usage, route optimization).
* **Expected / Desired Business Impact**:
  + Enhanced navigation experience and trust in Stellantis-brand vehicles.
  + Reinforced reputation for innovation and sustainability.

**3. Data Section**

* **Time Period**:
  + Historical sensor data from the past 6–12 months (initially).
  + Public automobile telematics dataset (Kaggle) for proof-of-concept.
* **Amount of Data**:
  + Large-scale time-series logs per vehicle (speed, location, energy usage).
  + Subset or pilot sample to manage complexity (e.g., limited geographic region, selected vehicle models).
* **Population**:
  + Vehicles equipped with relevant sensors (focus initially on newer models or a sample).
* **Labels**:
  + Driving patterns: derived or labeled from sensor traces (e.g., aggressive vs. cautious).
  + Trip outcomes: final ETA error, energy consumption metrics.
* **Train / Test Splits**:
  + Standard 80/20 or 70/30 split; cross-validation approach for robust performance estimates.
  + Potential separate hold-out for future validation.

**4. Solution Section**

* **Baseline Model**:
  + Stellantis’ current production models for ETA and energy consumption (serves as “no driving-style feature” baseline).
* **Initial Modeling Approach**:
  + **Feature Engineering**: Create driving-style features (acceleration variance, braking patterns, speed distribution, etc.).
  + **Classification**: Model the driving style as discrete categories (e.g., “aggressive,” “moderate,” “eco-friendly”).
  + **Integration**: Incorporate driving-style features (or predicted categories) into existing ETA and consumption models.
* **Assumptions & Risks**:
  + **Product / Data Assumptions**: Sensor data is reliable and sufficiently representative of real driving behavior.
  + **Risk**: Incomplete or noisy sensor data; real-world conditions differ from training data.
  + **Engineering Risk**: Integration complexities with existing systems.
  + **Business Risk**: Benefits might be less than expected if driving-style patterns don’t significantly affect predictions in some contexts.
* **Evaluation Metrics**:
  + **Classification**: Accuracy, F1-score, confusion matrix for driving-style classifier.
  + **ETA & Energy Predictions**: Reduction in MAPE (Mean Absolute Percentage Error), or RMSE for both ETA and energy usage.

**5. Requirements Section**

* **DataOps**:
  + Data ingestion pipeline for continuous sensor data.
  + Tools for data cleaning, missing data handling, outlier detection.
* **Labeling & Data Curation**:
  + Potential manual/heuristic labeling for driving style (if auto-labeled approach insufficient).
  + Processes to handle multi-source data (vehicle, environmental, public dataset).
* **DataDev Requirements**:
  + Automated jobs or scripts for feature generation (acceleration events, speed profiles).
  + Efficient storage structures for large time-series data.
* **Engineering Requirements**:
  + Model hosting environment (on-prem or cloud).
  + API endpoints for in-vehicle or telematics platform consumption.
  + Monitoring solution for model performance in production.
* **A/B Test**:
  + Potential pilot in a controlled environment or subset of vehicle fleets.
  + Compare new driving-style-enhanced predictions vs. baseline.

**6. Legal Consult Requirements**

* **Data Usage**:
  + Review GDPR implications if personal data (e.g., precise location + timestamps) is used.
  + Ensure compliance with Stellantis data privacy policies for in-vehicle data.
* **Open Source / Packages**:
  + Verify licensing for libraries used in modeling and data processing.
* **AI Regulations**:
  + Follow emerging automotive and data-protection regulations.
  + Maintain transparency in algorithmic decision-making if relevant (e.g., user consent, disclaimers).

1. **Intro section:**
   1. High level project description (TL;DR - goal, client, companies involved..)
   2. Timeline of major milestones: algo review, presentation to stakeholders, deployment, A/B test
   3. Team (data scientists, engineers, labelers, product owner…)

Introduction

* **High-Level Description**
  + **Project Title**: Enhancing ETA and Energy Consumption Predictions Through Driving Pattern Analysis