Project	Product Catalogue Standardization and Matching
Area	Data Management, Algorithm Development, User Interface (UI) and System Integration, and Validation + AI, ML, NLP & NLU
Duration	16 Weeks
Goal	The project aims to standardize and unify product data across multiple sources. By using AI and other algorithms, it will automatically identify and de-duplicate existing and future product entries. The final outcome is a single, clean catalogue that improves data integrity and enables seamless integration with other business systems.
Key Considerations	**Adaptable to new Requirements:** The project should use a modular design, making it easy to add new data sources or refine matching logic without rebuilding the entire system. **Simple to maintain and scale**: A well-documented architecture and a clear, automated monitoring dashboard will simplify maintenance and allow for seamless scaling as the product catalog grows. **Robust**: The solution must incorporate automated error handling, data validation checks, and a reliable data migration process to ensure data integrity and prevent system failures.

Weekly Milestones

Here is the project plan for Product Catalogue Standardization and Matching as a weekly plan over 4 months (16 weeks).

Week	Phase	Objective	Key Result
Week 1	Project Initiation & Scoping	Define the project's scope, goals, and team.	High-level plan & scope approved.
Week 2	Data Ingestion & Auditing	Ingest all product data into a staging environment.	All product data ingested and a data quality baseline established.
Week 3	Data Ingestion & Auditing (cont.)	Analyze the quality of the ingested data.	A detailed data profiling report is generated.
Week 4	Data Cleaning & Standardization	Begin the process of cleaning and normalizing attributes.	Initial set of data cleaning and normalization rules are defined and implemented on a sample dataset.
Week 5	Data Cleaning & Standardization (cont.)	Define and implement a master taxonomy for categorization.	A master list of product categories is created, and products are mapped to it.
Week 6	Develop Core De-duping Algorithms	Develop the core logic for rule-based and fuzzy matching.	A working prototype of the de-duplication engine is created.
Week 7	Develop Core De-duping Algorithms (cont.)	Integrate NLP, ML, and AI into the de-duping process.	The de-duplication engine can process natural language variations and abbreviations.
Week 8	Develop Core De-duping Algorithms (cont.)	Integrate NLP, ML, and AI into the de-duping process. contd.	contd...
Week 9	Build Review & Confirmation Interface	Develop and deploy a user-friendly interface for admin review.	A functional web interface for admin review is deployed in a test environment.
Week 10	Comprehensive Testing	Perform rigorous functional testing to ensure the matching engine accuracy is above a certain threashold. For any cases incorrectly matched, those should be handled manually using UI developed under - "Build Review & Confirmation Interface"	

Week 11	Pilot De-duplication Run & Validation	Run a pilot test and validate the accuracy of the algorithms.	A pilot de-duplication run is completed on a representative dataset, and a human-validated accuracy rate is established.
Week 12	Data Migration & Auxiliary Data Handling	Define and create a script for handling auxiliary/dependent data objects	A data migration strategy is defined, and a robust migration script is created and tested.
Week 13	Full-Scale De-duplication Run	Execute the de-duplication process on the entire product catalogue.	The full-scale de-duplication run is initiated.
Week 14	Full-Scale De-duplication Run (cont.)	Admin reviews and confirms potential duplicates.	A significant portion of potential duplicates have been reviewed and either confirmed or rejected by the admin.
Week 15	Monitoring & Continuous Improvement	Develop and implement a monitoring dashboard.	A monitoring dashboard is in place, and a process for handling new duplicates is defined.
Week 16	Monitoring & Continuous Improvement (cont.)	Finalize project, train users, and define a long-term maintenance plan.	User training is completed, and a long-term maintenance plan is in place.