**Project Roadmap**

**Phase 1: Foundation & Data Infrastructure (Weeks 1–8)**

**Week 1: Development Environment Setup Done**

* **Days 1–2**: Python environment, virtual environment, core packages
* **Days 3–4**: PostgreSQL database setup and configuration
* **Days 5–7**: Project structure, chemistry libraries, testing setup

**Week 2: Database Schema Design Done**

* **Days 1–2**: Core database models (Products, Companies, Relationships)
* **Days 3–4**: Chemical properties and user interaction tables
* **Day 5**: Database initialization and testing

**Week 3: Web Scraping Infrastructure Done**

* **Days 1–2**: Base scraper framework with rate limiting
* **Days 3–4**: PubChem and ChemSpider scrapers
* **Day 5**: Scraping manager and data storage integration

**Week 4: Data Collection & Quality**

* **Days 1–2**: Automated scraping pipelines
* **Days 3–4**: Data validation and cleaning processes
* **Day 5**: Initial chemical database population

**Week 5: Basic Chemical Intelligence**

* **Days 1–2**: Chemical name parsing and CAS number validation
* **Days 3–4**: Basic chemical property assignment
* **Day 5**: Product categorization and family creation

**Week 6: Infrastructure Completion**

* **Days 1–2**: Logging, monitoring, and error handling
* **Days 3–4**: Backup systems and data recovery
* **Day 5**: Performance optimization and indexing

**Week 7: OCR & PDF Processing Setup**

* **Days 1–2**: Tesseract OCR installation and configuration
* **Days 3–4**: PDF text extraction pipeline
* **Day 5**: Chemical entity extraction from documents

**Week 8: NLP Processing**

* **Days 1–2**: spaCy setup for chemical name recognition
* **Days 3–4**: Context-aware chemical extraction
* **Day 5**: Confidence scoring and validation

**Phase 2: Machine Learning Pipeline (Weeks 9–12)**

**Week 9: ML Environment & Data Preparation**

* **Days 1–2**: TensorFlow, PyTorch, scikit-learn setup
* **Days 3–4**: Feature engineering pipeline for chemical data
* **Day 5**: Data quality analysis and validation

**Week 10: Property Prediction Models**

* **Days 1–2**: Regression models for physical properties
* **Days 3–4**: Neural networks for complex property relationships
* **Day 5**: Model evaluation and hyperparameter tuning

**Week 11: Chemical Similarity Engine**

* **Days 1–2**: RDKit molecular fingerprinting
* **Days 3–4**: Tanimoto similarity calculations
* **Day 5**: Similarity matrix generation and storage

**Week 12: Recommendation System**

* **Days 1–2**: Collaborative filtering implementation
* **Days 3–4**: Content-based recommendations using similarity
* **Day 5**: Hybrid recommendation system

**Phase 3: User Intelligence & Recommendations (Weeks 13–18)**

**Week 13: User Behavior Tracking**

* **Days 1–2**: Website analytics integration
* **Days 3–4**: Custom event tracking for chemical interactions
* **Day 5**: Real-time user session analysis

**Week 14: Customer Analytics**

* **Days 1–2**: User segmentation and profiling
* **Days 3–4**: Purchase pattern analysis
* **Day 5**: Customer lifetime value prediction

**Week 15: Advanced Recommendations**

* **Days 1–2**: Personalized recommendation algorithms
* **Days 3–4**: Cross-selling optimization
* **Day 5**: Recommendation explanation system

**Week 16: Business Intelligence**

* **Days 1–2**: Sales analytics and reporting
* **Days 3–4**: Chemical demand forecasting
* **Day 5**: Market trend analysis

**Week 17: Patent Monitoring System**

* **Days 1–2**: USPTO and Google Patents API integration
* **Days 3–4**: Patent text analysis and chemical extraction
* **Day 5**: Competitive intelligence dashboard

**Week 18: Competitive Analysis**

* **Days 1–2**: Market analysis and competitor tracking
* **Days 3–4**: New product opportunity identification
* **Day 5**: Patent landscape visualization

**Phase 4: Integration & Deployment (Weeks 19–24)**

**Week 19: API Development**

* **Days 1–2**: FastAPI framework setup and core endpoints
* **Days 3–4**: Authentication and rate limiting
* **Day 5**: API documentation and testing

**Week 20: API Completion**

* **Days 1–2**: Recommendation and similarity API endpoints
* **Days 3–4**: Batch processing APIs
* **Day 5**: Webhook system for real-time updates

**Week 21: Frontend Dashboard**

* **Days 1–2**: Admin dashboard for data management
* **Days 3–4**: Analytics visualization interface
* **Day 5**: User management and configuration UI

**Week 22: Customer-Facing Interface**

* **Days 1–2**: Chemical search and filtering interface
* **Days 3–4**: Recommendation display system
* **Day 5**: Chemical compatibility checker

**Week 23: Testing & Quality Assurance**

* **Days 1–2**: Unit testing for all components
* **Days 3–4**: Integration and end-to-end testing
* **Day 5**: Load testing and performance optimization

**Week 24: Production Deployment**

* **Days 1–2**: Production infrastructure setup
* **Days 3–4**: Deployment automation and monitoring
* **Day 5**: Launch, documentation, and handover