

# **Product Requirements Document (PRD)**

## **Intelligent Document Classifier and Router**

### **1. Executive Summary**

#### 1.1 Product Vision

Automate and optimize document management by leveraging AI to classify, route, and organize documents intelligently, reducing manual effort and improving productivity.

#### 1.2 Business Objectives

- Productivity: Save 1 workday/week per knowledge worker
- Speed: Reduce routing time by 70% (from 3 days to 8 hours)
- Accuracy: Achieve 95% classification accuracy
- Searchability: Cut document search time by 80%

#### 1.3 Target Market

- Legal, consulting, finance sectors
- Document-heavy workflows
- Compliance-driven organizations
- Remote/hybrid teams

### **2. Product Overview**

#### 2.1 Core Value Proposition

AI-powered document intelligence that delivers context-aware classification, routing, and actionable insights.

#### 2.2 Key Features

- Intelligent Classification
- Smart Routing
- Content Extraction
- Workflow Integration
- Knowledge Discovery

### **3. Architecture Overview**

#### 3.1 Microservices

- API Gateway (FastAPI)
- Classification Service (spaCy, Python)
- Routing Engine (Rule Engine)
- Content Analysis (Pandas, NumPy)
- AI/ML Service (scikit-learn)
- Workflow Integration (Slack, Jira, Email)
- Search & Discovery (SQLAlchemy, Full-Text Search)

### 3.2 Data Layer

- Document Storage (MinIO/S3)
- Metadata DB (PostgreSQL)
- Permissions DB

### 3.3 External Integrations

- OpenAI API
- Slack/Microsoft Teams
- Email Services

## **4. Feature Specifications**

### 4.1 Document Classification

- Multi-format support (PDF, DOC, TXT, images)
- AI-powered classification with confidence scoring
- Entity extraction, sentiment analysis, topic modeling

### 4.2 Intelligent Routing

- Role-based and expertise-based routing
- Business rule engine with escalation and approval workflows
- Context-aware processing (project, customer, compliance)

### 4.3 Workflow Integration

- Slack, Teams, Email, Jira, Asana, Trello
- Calendar and mobile push notifications
- Custom webhooks

### 4.4 Search & Discovery

- Full-text and semantic search

- Faceted filters and auto-complete
- Knowledge graphs and recommendation engine

## **5. Technical Requirements**

### 5.1 Performance

- <10s per document
- 500+ docs/hour throughput
- <1s search response
- 99.5% uptime
- 200+ concurrent users

### 5.2 AI/ML

- 95% classification accuracy
- Continuous learning
- Multi-language support
- Custom model training

### 5.3 Integration

- RESTful APIs
- WebSocket for real-time updates
- SSO (OAuth2, SAML)
- Data export formats

## **6. User Experience**

### 6.1 Web Dashboard

- Document pipeline visualization
- Quick actions and analytics
- Rule and workflow configuration

### 6.2 Mobile Interface

- Camera-based upload
- Offline capabilities
- Notification center

### 6.3 Admin Console

- Role and permission management
- Rule configuration
- Model training and analytics

## **7. Security & Compliance**

- End-to-end encryption
- Granular access control
- Audit logging
- GDPR, HIPAA, SOX compliance
- Data retention policies

## **8. Success Metrics**

### 8.1 Operational

- Classification accuracy
- Processing time
- User adoption
- Search success rate

### 8.2 Business Impact

- Time savings
- Productivity gains
- Error reduction
- User satisfaction (NPS)

## **9. Implementation Roadmap**

### Phase 1: Core Classification (2 Days POC)

- Setup project structure, Docker, DB schema
- Build API Gateway and Classification MVP
- Integrate basic routing and notifications

### Phase 2: Enhanced Intelligence (Week 1–2)

- Advanced classification and routing
- Slack/Teams integration

- Semantic search and analytics

#### Phase 3: Enterprise Features (Week 3–4)

- Multi-tenancy
- Advanced security and compliance
- BI dashboards

#### Phase 4: AI Enhancement (Month 2)

- Predictive routing
- Knowledge graphs
- Continuous learning and optimization

### **10. Risks & Mitigation**

- AI Accuracy: Human-in-the-loop review
- Scalability: Microservice architecture
- Integration Complexity: API-first design
- Data Privacy: Privacy-by-design principles

### **11. Competitive Advantages**

#### 11.1 Technical

- Native AI-first architecture
- Real-time processing
- Context-aware intelligence
- Continuous learning

#### 11.2 Business

- Rapid deployment (2-day POC)
- Cost-effective (80% cheaper than competitors)
- Intuitive UI
- Flexible integrations

### **12. Conclusion**

The Intelligent Document Classifier and Router is a transformative solution for modern businesses, combining AI, microservices, and seamless integrations to deliver scalable, secure, and intelligent document management.