IDP- Developing a PCF/CCF AI Tool at Parabella Analytics

03.06.2024

Project: Development of an AI-enhanced Software module for Product Carbon Footprint (PCF) and Corporate Carbon Footprint (CCF) Calculation for ESG Reporting enhanced with AI

1. Motivation and Background

Environmental, Social, and Governance (ESG) data collection and reporting are rapidly gaining importance for companies. Specifically, calculating the Product Carbon Footprint (PCF) and Corporate Carbon Footprint (CCF) is fundamental for sustainability strategies and regulatory compliance (e.g., CSRD). Current methods are often manual, time-consuming, data-intensive, and prone to errors, especially with complex value chains and Scope 3 emissions. Parabella Analytics, an AI-driven ESG analytics startup, aims to address these challenges through innovative software solutions. This IDP will significantly contribute to developing two core software modules: a PCF tool and a CCF tool, both enhanced with Artificial Intelligence (AI) to automate processes, increase efficiency, and improve accuracy.

2. Project Goals and Interdisciplinarity

The primary goal is the design and prototypical development of an AI-enhanced software tool to assist companies in calculating and analyzing their PCF and CCF. Specific objectives include:

- **PCF Module Development:** Calculate greenhouse gas emissions of a product across its lifecycle (Cradle-to-Gate or Cradle-to-Grave). Utilize AI for intelligent emission factor selection, data gap filling, and uncertainty analysis.

- **CCF Module Development:** Capture and calculate a company's direct and indirect emissions according to common standards (e.g., GHG Protocol for Scope 1, 2, and 3). Employ AI for improved estimation of Scope 3 emissions (e.g., from purchased goods/services, transportation).
- **AI Integration:** Investigate and implement suitable AI methods (e.g., Natural Language Processing (NLP) and GenAI for data extraction from supplier documents, GenAI for predictive models and anomaly detection). (Some software modules already exist can be partially used)
- **User Experience:** Develop an intuitive User Interface (UI) for data input, configuration, and visualization of results.

Interdisciplinarity: The project merges core Informatics competencies (software development, data management, AI/ML) with expertise from the application subject (Strategy and Organization, Business Informatics, Sustainability Management). Students will apply informatics methods to a real-world business problem in the rapidly growing ESG sector, considering strategic implications for companies.

Moreover, an Excel-based prototype is already in place, and a set of architecture diagrams is available to give us a solid head start.

3. Planned Methodology (Scrum) and Technology Stack (Proposal)

The project will follow the **Scrum agile development framework** to ensure iterative progress, flexibility, and continuous feedback. The project duration is 3 months (full-time).

- **Scrum Roles:**
    - **Product Owner:** Domink Erb and Alexander Pavlovski. Responsible for defining the product vision, managing and prioritizing the Product Backlog.
    - **Scrum Master:** One of the students (can rotate) or the academic supervisor if they are actively involved. Facilitates the Scrum process, removes impediments.
    - **Development Team:** The 3-4 students. Self-organizing team responsible for delivering a potentially shippable increment of the product at the end of each Sprint.
- **Scrum Events:**

    - **Sprints:** Fixed-length iterations of **2 weeks**. (For a 12-week project, this means 6 Sprints).
    - **Sprint Planning:** At the beginning of each Sprint to define the Sprint Goal and select items from the Product Backlog for the Sprint Backlog.
    - **Daily Scrum:** Short daily meeting to synchronize activities and create a plan for the next 24 hours.
    - **Sprint Review:** At the end of each Sprint to inspect the increment and adapt the Product Backlog if needed. Stakeholders (Parabella Analytics) will participate.
    - **Sprint Retrospective:** After the Sprint Review and prior to the next Sprint Planning, to reflect on the past Sprint and identify improvements for the next.
- **Scrum Artifacts:**
    - **Product Backlog:** An ordered list of everything known to be needed in the product. Managed by the Product Owner.
    - **Sprint Backlog:** The set of Product Backlog items selected for the Sprint, plus a plan for delivering the product Increment and realizing the Sprint Goal.
    - **Increment:** The sum of all Product Backlog items completed during a Sprint and all previous Sprints. A "Done" increment is usable and potentially releasable.
- **Definition of Done (DoD):** A shared understanding of what it means for work to be complete, ensuring quality. To be defined by the team at the start.
- **Technology Stack (Proposal):**
    - **Backend:** Python with FastAPI
    - **Database:** PostgreSQL with vector extension
    - **Frontend:** React with Typescript
    - **AI/ML:** Python libraries (scikit-learn, TensorFlow/PyTorch, spaCy/NLTK etc.
    - **Version Control:** GitLab
    - **Project Management:** JIRA + GitLab

4. Expected Results and Deliverables

- A functional software module of the PCF and CCF tools with core AI-enhanced features embedded in Parabellas software landscape

- A well-maintained Product Backlog and Sprint Backlogs.
- Regularly demonstrated increments at Sprint Reviews.
- Comprehensive technical documentation (architecture, data models, API descriptions).
- A detailed final project report outlining the process, Sprints, results, and limitations.
- A final presentation/demonstration of the project and its outcomes.
- Source code/gitlab branch of the developed module.

5. Task Distribution within the Team (Self-Organizing Scrum Team)

While students may have individual strengths (e.g., one focuses more on AI/ML, another on backend, another on frontend/UI, and one on ESG domain integration & testing), the Development Team is cross-functional and self-organizing. They will collectively decide how to best accomplish the work in each Sprint Backlog. All team members are jointly responsible for project planning, documentation, and delivering the Sprint Goals.

6. Timeline and Milestones (Based on 6 x 2-Week Sprints over 3 Months)

- **Sprint 0 (Part of Week 1, if needed, or integrated into Sprint 1):**
  - Team formation, role clarification (Scrum Master initial assignment).
  - Detailed introduction to Parabella Analytics, PCF/CCF standards, AI application possibilities.
  - Confirmation of accompanying lecture.
  - Initial Product Backlog creation and refinement with the Product Owner.
  - Setup of development environment, tools (Git, JIRA/Trello). Define initial DoD.
  - **Goal:** Project kick-off, initial Product Backlog ready, tech stack confirmed, Sprint 1 planning prepared.
- **Sprint 1 (Weeks 1-2):**
  - **Sprint Goal Example:** "Develop a basic PCF data model and calculation logic for a single product with manual data input and placeholder emission factors, accessible via a rudimentary UI."

    - **Activities:** Sprint Planning, implement core PCF backend logic, database schema for PCF, basic input form. Daily Scrums. Sprint Review & Retrospective.
- **Sprint 2 (Weeks 3-4):**
    - **Sprint Goal Example:** "Enhance PCF module with AI-driven emission factor suggestion for one category and improve PCF UI for better data entry and result display."
    - **Activities:** Research/implement first AI component for PCF, refine PCF UI, unit tests.
- **Sprint 3 (Weeks 5-6):**
    - **Sprint Goal Example:** "Develop a basic CCF data model (Scope 1 & 2) and calculation logic with manual input, accessible via a new rudimentary UI section."
    - **Activities:** Implement core CCF backend logic, database schema for CCF, basic CCF input form.
- **Sprint 4 (Weeks 7-8):**
    - **Sprint Goal Example:** "Enhance CCF module with AI-assisted estimation for one Scope 3 category (e.g., purchased goods) and integrate CCF UI with the existing framework."
    - **Activities:** Research/implement first AI component for CCF, refine CCF UI, start considering integration points.
- **Sprint 5 (Weeks 9-10):**
    - **Sprint Goal Example:** "Integrate PCF and CCF modules into a unified prototype with a basic dashboard showing key metrics for both, and conduct initial user feedback sessions with Parabella Analytics."
    - **Activities:** Focus on module integration, develop dashboard components, prepare for and conduct UAT.
- **Sprint 6 (Weeks 11-12):**
    - **Sprint Goal Example:** "Deliver a polished and tested prototype addressing feedback from Sprint 5, complete core documentation, and prepare the final project presentation."
    - **Activities:** Bug fixing, performance optimization, final UI/UX polish, complete technical documentation, prepare final report outline, prepare presentation.
- **Post-Project (Week 13, if buffer needed or for final wrap-up):**
    - Finalize project report and all deliverables.
    - Official project handover to Parabella Analytics.
    - Final project presentation to Examiner and stakeholders.

8. Supervision and Resources

- **Product Ownership & Industry Supervision:** By Dominik Erb and Alexander Pavlovski regarding ESG domain knowledge, application requirements, Product Backlog prioritization, and feedback on Sprint Increments.
- **Academic Supervision:** By the Examiner at TUM.
- **Resources from Parabella Analytics:** Provision of domain expertise, (anonymized) sample datasets if available and legally permissible (otherwise work with synthetic or public data), access to relevant standards and literature. Regular availability for Sprint Reviews and backlog refinement.
