

Project Number: 17

Project Title: Statsplainer: An online reader application to understand statistics in academic literature

Project Clients: Main: Benjamin Tag, Co: Simo Hosio

Project specializations: Human Computer Interaction (HCI);Artificial Intelligence (Machine/Deep Learning, NLP);Web Application Development;

Number of groups: 3 groups

Main contact: Main: Benjamin Tag, Co: Simo Hosio

Background:

Academic papers, industry reports, and technical documents often contain complex statistical data that can be difficult to interpret for non-experts. Understanding statistical concepts is crucial in fields such as healthcare, business, policy-making, and research, but many readers struggle to comprehend the meaning and implications of statistical results.

Existing solutions like statistical glossaries and educational videos provide static explanations, but they lack interactivity and contextual relevance. Recent research in Human-Computer Interaction (HCI) and AI has explored ways to simplify complex data through interactive explanations.

Project Goal:

The goal of StatSimplify is to build an interactive online PDF reader that allows users to:

- ✓ Upload a PDF document
- ✓ Highlight a paragraph containing statistical information
- ✓ Automatically generate intuitive explanations in a sidebar, including:
 - A simplified breakdown of the statistical terms
 - A real-world analogy for better understanding
 - An “Explain Like I’m 5” (ELI5) explanation of its significance and application

This project will make statistical content more accessible, interpretable, and user-friendly, particularly for students, researchers, and professionals who may not have a strong background in statistics.

Requirements and Scope:

The project will focus on the following core components:

In-Scope:

- PDF Upload & Processing – Users can upload PDF documents and extract text
- Statistical Text Recognition – Identify statistical terms, numbers, and concepts within a highlighted paragraph
- Automatic Explanation Generation – Provide explanations, real-world analogies, and ELI5 descriptions
- Interactive UI with Sidebar Display – Users can highlight a paragraph, triggering an interactive sidebar with explanations
- Expandable Knowledge Base – The system should improve over time by incorporating additional statistical explanations
- Web-Based Deployment – The tool will be accessible via a web application

Additional, if time allows:

- Full AI-powered explanation generation (Instead, we will use predefined templates and rule-based methods)

Required Knowledge and skills:

Detailed Features & Specifications:

1. PDF Processing & Text Extraction

- Users can upload PDFs
- Extract and display text content
- Highlighting functionality for users to select specific paragraphs

2. Statistical Concept Recognition

- Identify statistical terms and values (e.g., p-values, confidence intervals, regression coefficients)
- Extract and categorize statistical terms from text using Natural Language Processing (NLP)
- Use open-source repositories like Statslator.js and Charagraph

3. Explanation Generation Engine

- Provide three levels of explanation:
 1. Basic definition of the statistical concept

2. Real-world analogy for better understanding
3. “Explain Like I’m 5” (ELI5) explanation
 - Develop a predefined knowledge base of statistical concepts and explanations
 - (Optional) Experiment with AI models to enhance explanation quality
4. Interactive UI with Sidebar Display
 - Users can highlight a paragraph
 - Sidebar dynamically provides:
 - Recognized statistical terms
 - Corresponding explanations and analogies
 - UI should be intuitive and minimalistic for easy interaction
5. Web-Based Application
 - Develop a responsive web app using React.js or Vue.js
 - Backend API to handle text processing and explanations
 - Deploy using AWS, GCP, or Firebase

Expected outcomes/deliverables:

=> A functional web-based prototype, including:

- PDF Upload & Highlighting Feature
- Explanation Sidebar with statistical concept breakdown
- Interactive UI for easy user engagement

=> Codebase (open-source) hosted on GitHub

=> Technical Documentation

- Architecture and implementation details
- API documentation

=> User Guide & Demo Video

- Instructions on how to use the tool