The Use of Generative AI to generate estimates for the project plan, design document, source code, test plan and test cases, user API documentation.

Group 17
Soham Kamble (A20517098)
Bhavya Chawla (A20516957)

Content:

- **Project Overview**
- **Functional, Non Functional Requirements**
- **Tools**
- **Data Sets**

Project Overview

- **What is Generative AI?**

  Generative AI refers to a cutting-edge technology that can create content, such as text or code, using machine learning algorithms. It has the potential to revolutionize the estimation process in project management.

- **Benefits of using generative AI for estimates:**

  Generative AI streamlines the estimation process by automating repetitive tasks, reducing human bias, and improving accuracy, ultimately leading to more reliable project plans and documents.

- **Challenges of using generative AI for estimates:**

  Despite its advantages, using generative AI for estimates presents challenges like data quality, model training, and ethical concerns.

Functional and Non Functional Requirements

Functional	Non Functional
Task-Specific Estimation	Accuracy of Estimates: The AI system's project plan and cost estimates should be accurate.
Customization	Cost Control: The project should stay within budget.
Real-time Estimation	Response Time: The AI system should respond within seconds
Data Integration	Scalability: The AI system should handle varying workloads.
Validation and Verification	Data Security: The AI system should ensure data security and privacy.

Tools

Data Set

The SEERA Software Cost Estimation Dataset - https://zenodo.org/records/3987969

- It is the result of the collection of 120 software development project data from 42 organizations and dataset contains 76 attributes
- The years taken in consideration for the data 1993 to 2019
- Total of actual efforts 1087311.5 person hours
- Total of estimated efforts 627022 person hours

Dataset Format	• The dataset is in Excel format. • Column headings represent attributes/sub-attributes. • Rows represent projects. • Missing values are represented by "?" • The names of sub-attributes not included in the formulas of their main attributes are preceded with a (-). • For values that depend on the answers of conditional questions: if the condition is not met the value is encoded as "N/A" and is not considered a missing value.
Information included within the dataset	• ID of organizations that contributed data • Identifiers of projects in each group • Customers type • Start/end date of projects
Dataset information included in research paper	• Attributes with outliers • Method/tool used to identify outliers • Number and proportion of incomplete data • List of problems encountered during data collection • Redundant data and reason for redundancy

Effort	
Estimated effort	Actual effort
2112	3168
1056	1584
3168	5280
5280	5280
19008	38016
7392	9856
5280	7920
4400	4400
4224	4224
6468	24255

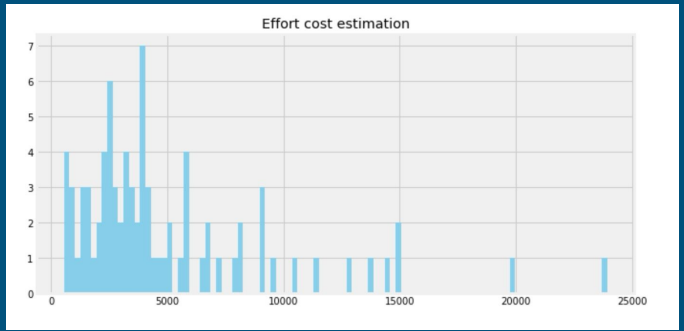
Size		
Object points	Other sizing meth	Estimated size
100	2	N/A
19	2	N/A
64	3	50
246	5	70
420	5	120
177	5	76
54	5	5
71	5	50
14	5	5
335	4	6

Data Set

Software Cost estimation - https://www.kaggle.com/code/raulvilla/sofware-cost-estimation-i

- It is the result of the collection of 81 software development project
- The years taken in consideration for the data 1982 to 1988
- Total of actual efforts 408751 person hours

	id	Project	TeamExp	ManagerExp	YearEnd	Length	Effort	Transactions	Entities	PointsNonAdjust	Adjustment	PointsAjust	Language
0	1	1	1	4	85	12	5152	253	52	305	34	302	1
1	2	2	0	0	86	4	5635	197	124	321	33	315	1
2	3	3	4	4	85	1	805	40	60	100	18	83	1
3	4	4	0	0	86	5	3829	200	119	319	30	303	1
4	5	5	0	0	86	4	2149	140	94	234	24	208	1
5	6	6	0	0	86	4	2821	97	89	186	38	192	1
6	7	7	2	1	85	9	2569	119	42	161	25	145	2
7	8	8	1	2	83	13	3913	186	52	238	25	214	1
8	9	9	3	1	85	12	7854	172	88	260	30	247	1
9	10	10	3	4	83	4	2422	78	38	116	24	103	1



Data Set - others

- PROMISE Software Engineering Repository - http://promise.site.uottawa.ca/SERepository/index.html

  A collection of publicly available datasets and tools to serve researchers in building predictive software models (PSMs)

- International Software Benchmarking Standards Group - https://www.isbsg.org/
- GitHub
- Kaggle Datasets

Project Timeline

- **Week 1**: Project Initiation, Data Collection & Setup - Completed
- **Week 2**: Model Training, Document Generation & Preliminary Review - In Progress
- **Week 3**: Review Corrections, Finalization & Reporting

Thank you

- https://platform.openai.com/docs/guides/gpt
- https://www.nvidia.com/en-us/glossary/data-science/generative-ai/
- https://www.linkedin.com/pulse/how-generative-ai-can-help-you-improve-employee-support-documentation#:~:text=API%20Documentation,for%20developers%20and%20the%20community.