

BullSolve-A-Mathematical-Problem-Solving-Model

BullSolve is an AI-powered system designed to solve mathematical problems through a dual methodology approach. It combines:

- A shallow neural network that classifies mathematical problems by type
- A fine-tuned large language model (LLaMA-3.1-8B) that generates step-by-step solutions

The system can analyze math problems across various domains including algebra, geometry, calculus, and probability. It not only provides answers but explains the reasoning process, making it valuable for educational purposes.

Technical Description

- **First Approach:** A shallow neural network architecture that:
 - Processes text and equations via embedding layers
 - Uses multilayer perceptron for classification
 - Achieves 85% accuracy in problem type classification
 - Employs different representations (BOW, TF-IDF, Word2Vec)
- **Second Approach:** A fine-tuned language model that:
 - Uses a 4-bit quantized version of LLaMA-3.1-8B
 - Adapts using Low-Rank Adaptation (LoRA)
 - Generates detailed step-by-step solutions
 - Employs budget forcing to prevent infinite reasoning loops
- **Data Processing:**
 - Uses the MATH dataset with 7 mathematical categories
 - Employs Math-Augmented dataset with 1,006 validated algebra problems
 - Implements data augmentation using large language models
- **Key Features:**
 - Problem type classification
 - Step-by-step solution generation with explanations
 - Transparent reasoning process
 - Educational value through detailed explanations

Results

Neural Network Classifier

- **Best Overall Accuracy:** 85% with optimized MLP architecture
- **Performance by Problem Type:**
 - Algebra: 91% accuracy (merged from Algebra, PreAlgebra, and Intermediate Algebra)
 - Number Theory: 90% accuracy
 - Lower performance on underrepresented classes

Language Model Solution Generator

- **First Evaluation Method** (Problem statement only):
 - Accuracy: 33.3%
 - ROUGE-1 F1: 0.337
 - ROUGE-2 F1: 0.173
 - ROUGE-L F1: 0.319
- **Second Evaluation Method** (With prompt template):
 - Accuracy: 100%
 - ROUGE-1 F1: 0.411
 - ROUGE-2 F1: 0.208
 - ROUGE-L F1: 0.393
- **Comparative Analysis:**
 - Traditional classifiers (baseline): 33-65% accuracy
 - Shallow neural networks: 70-75% accuracy
 - Fine-tuned architecture: 85% accuracy
 - Significant improvement using prompt templates for solution generation

Prototype

video: <https://youtu.be/bE5qk9Pz1jo>

The BullSolve model is available for testing on Hugging Face. You can access the model through the following resources:

- [BullSolve Model](#)
- [Math-augmented-dataset](#)
- [Colab](#)

Project Details

This project was developed by:

- Kevin Fabio Ramos Lopez
- Kevin Camilo Rincon Bohorquez
- Nolhan Dumoulin

The model achieves promising results in mathematical problem solving, with particular strength in generating detailed, step-by-step solutions that offer educational value beyond just providing answers.