Thesis Proposal: AutoMMM - An AI-Powered System for Market Mix Modeling

Prateek Nigam

June 2025

Abstract

This thesis will develop AutoMMM, an AI-powered system to analyze marketing data, like ad spend and sales, automatically. Using Python, LangChain, and LangGraph, AutoMMM will use multiple AI agents to process data, run models, and suggest better marketing strategies. It will generate synthetic data to test the system and validate its performance. This report explains the problem, goals, methods, expected results, and a two-month plan.

1 Introduction

Market Mix Modeling (MMM) helps businesses see how marketing efforts, like ads and promotions, drive sales. It guides smart budget decisions to maximize return on investment (ROI) [1]. Traditional MMM is slow and needs a lot of manual work, which struggles in today's fast digital marketing world [2].

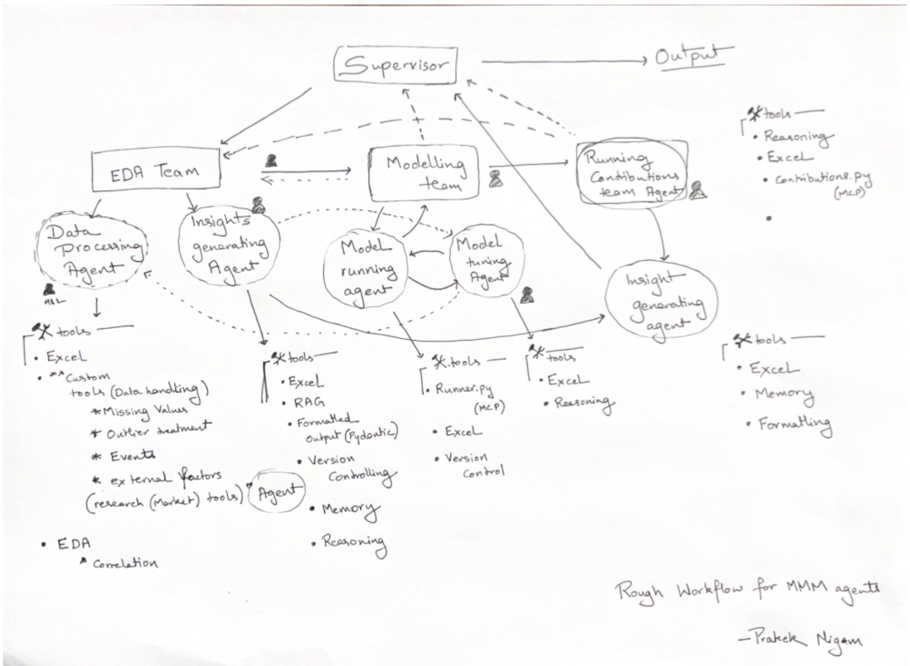
Artificial Intelligence (AI) can improve MMM by automating tasks and finding data patterns [3]. However, there are few systems that fully automate MMM using multiple AI agents. AutoMMM will solve this by using Python, LangChain, and LangGraph to analyze ad spend, sales, and external factors, delivering clear business advice.

The problem is the lack of fully automated MMM systems with multi-agent collaboration. The goals are to:

1. Build an automated MMM system with AI agents.
2. Test it using synthetic data and validate performance.
3. Add new ideas to AI in marketing analytics.

2 Literature Review

MMM uses math to measure marketing's impact on sales [4]. AI makes MMM faster by handling big data and giving quick insights [5]. Tools like Adobe Mix Modeler [6] and NextBrain AI [7] use AI but still need human help for some tasks [8]. Multi-agent systems (MAS) let AI agents share tasks, but they're rarely used in MMM [9]. AutoMMM will fill these gaps by automating MMM and using MAS for teamwork.



3 Methodology

3.1 System Design

AutoMMM will be a multi-agent system managed by a **Supervisor Agent** using LangGraph for workflow orchestration and LangChain for agent communication. The system will include:

- **Data Team**:
  - **Data Processing Agent**: Will run Python scripts for exploratory data analysis (EDA) on ad spend, sales, and external factors, creating Excel reports with charts and notes on data issues like missing values or outliers.
  - **Insight Agent**: Will summarize key performance indicators (KPIs) and data stories using LangChain to guide modeling.
- **Modeling Team**:
  - **Model Running Agent**: Will run a simple regression model (MCP) to predict sales, saving model summaries and graphs (PNG files).
  - **Model Tuning Agent**: Will adjust hyperparameters using insights from the Data Team to improve accuracy.
- **Contribution Agent**: Will run `run_contributions.py` to calculate marketing contributions using the Shapley method, producing an Excel file.
- **Final Suggestions**: Will turn results into clear business advice, like optimizing ad spend for better ROI.

3.2 Synthetic Data Generation

To test AutoMMM, I will generate synthetic data that mimics real marketing data. The process will include:

- **Data Structure**: Create 104 weeks (2 years) of weekly data for ad spend (e.g., TV, digital), sales, and external factors (e.g., holidays, seasonality).
- **Realistic Patterns**: Add positive correlations between ad spend and sales, diminishing returns (saturation), and carryover effects (adstock) using decaying functions and logistic transformations.
- **Noise and Outliers**: Include Gaussian noise and occasional outliers to simulate real-world data challenges.
- **Ground Truth**: Set known coefficients (e.g., 0.3 for TV, 0.4 for digital) to compare with model outputs.
- **Config File**: Use a YAML config to define parameters like adstock decay, saturation, and noise levels for reproducibility.

The Data Processing Agent will handle this synthetic data, ensuring it's ready for modeling.

3.3 Model Validation

To check AutoMMM's performance, I will:

- **Compare Coefficients**: Compare model-estimated coefficients to ground truth coefficients from the synthetic data to measure accuracy.
- **Evaluate Metrics**: Use metrics like Mean Absolute Error (MAE) and R-squared to assess prediction quality.
- **Test Robustness**: Validate the model on noisy data and outliers to ensure it handles real-world challenges.
- **ROAS Analysis**: Check if the Contribution Agent's outputs align with expected return on ad spend (ROAS) based on ground truth.

The Supervisor Agent will ensure all agents work together to validate results.

3.4 Implementation Details

- **Tools**: Python for coding, LangChain for agent logic, LangGraph for workflows.
- **Modeling**: Use a simple regression model (MCP) for sales prediction.
- **Contributions**: Calculate using `run_contributions.py` with the Shapley method.
- **Data Generation**: Create synthetic data with realistic MMM patterns.

Table 1: AutoMMM Agent Jobs

Agent	Job
Supervisor Agent	Manages agents using LangGraph
Data Processing Agent	Analyzes data, generates Excel reports
Insight Agent	Summarizes KPIs using LangChain
Model Running Agent	Runs MCP model, saves results
Model Tuning Agent	Improves model settings
Contribution Agent	Calculates contributions with `run_contributions.py`
Final Suggestions	Gives business advice

4 Expected Results

4.1 Academic Results

- Develop a new way to automate MMM with AI agents.
- Create a framework for multi-agent systems in marketing analytics.

4.2 Practical Results

- Build a tool that automates MMM, saving time.
- Provide clear advice to improve marketing and boost profits.

5 Timeline

The project will take two months:

- **Week 1**: Study past work and design system.
- **Weeks 2-3**: Build Data Team and synthetic data generator.
- **Weeks 4-5**: Develop Modeling Team.
- **Week 6**: Create Contribution Agent.
- **Week 7**: Test system and validate performance.
- **Week 8**: Finalize testing and write report.

6 References

References

References

[1] LatentView, A complete guide to Marketing Mix Modeling, https://www.latentview.com/marketing-mix-modeling/, 2024.

Table 2: Project Timeline

Task	**Timeline**
Study and System Design	Week 1
Data Team and Synthetic Data	Weeks 2-3
Modeling Team	Weeks 4-5
Contribution Agent	Week 6
Testing and Validation	Week 7
Final Testing and Report	Week 8

[2] Ruler Analytics, What is Marketing Mix Modeling?, https://www.ruleranalytics.com/blog/analytics/marketing-mix-modelling/, 2024.

[3] DiGGrowth, The Role of AI in Marketing Mix Modeling, https://diggrowth.com/blogs/marketing-mix-modeling/the-role-of-ai-in-marketing-mix-modeling/, 2024.

[4] Ridhima Kumar, Market Mix Modeling (MMM) — 101, https://medium.com/data-science/market-mix-modeling-mmm-101-3d094df976f9, 2019.

[5] Pecan AI, How Machine Learning Speeds Marketing Mix Modeling, https://www.pecan.ai/blog/marketing-mix-modeling-machine-learning/, 2023.

[6] Adobe, Adobe Mix Modeler, https://business.adobe.com/products/experience-platform/adobe-mix-modeler.html, 2024.

[7] NextBrain AI, Marketing Mix Modeling, https://nextbrain.ai/marketing-mix-modeling, 2023.

[8] Rockerbox, How Can AI Help My Marketing Mix Modeling?, https://www.rockerbox.com/blog/how-can-ai-help-my-marketing-mix-modeling, 2024.

[9] Analytic Edge, Automation and Marketing Mix Modelling, https://analytic-edge.com/blog/automation-and-marketing-mix-modelling-balancing-ai-with-hu
2023.