

VERIFACT

By : Hardi Dave, Dhairya Patel, and Avi Patel

INTRODUCTION

- **This capstone project focuses on leveraging machine learning techniques to analyze the authenticity of news shared on social media platforms.**
- **Using Kaggle dataset to train and test the models identifying Positive & Negative News.**
- **This project addresses the growing spread of fake news in the digital age by tackling societal challenges of confusion, mistrust, and harm caused by false information.**

PROBLEM

- **Persona: Mark, 28, Social Media Influencer**
- **Quote: "As a social media influencer with a large following, I can't afford to accidentally share misinformation. I need a reliable way to quickly verify the authenticity of news articles and content before posting, so I can maintain trust and credibility with my audience."**
- **Key Points:**
 - **Time-consuming manual fact-checking**
 - **Risk of damaging credibility and trust with his large social media followers**
 - **Overwhelming volume of information to verify**

SOLUTION

- **For Mark:**
 - **AI-powered news and content authenticity checker**
 - **Quick analysis of articles and social media posts**
 - **Detailed reports with credibility scores and explanations**

UNIQUE VALUE PROPOSITION

- **Key differentiators:**
 - **User-friendly interface for quick checks**
 - **Detailed analysis reports for in-depth understanding**
 - **Continuous learning from user feedback**
 - **Collaborative flagging system**

ADDITIONAL PERSONAS

Anna, 45, Concerned Parent

- **Wants to ensure his children access reliable information online**

Emma, 32, Social Media Manager

- **Needs to verify content before sharing on company platforms**

Prof. Johnson, 55, University Lecturer

- **Encourages students to fact-check their sources**

MVP

- **AI-powered authenticity analysis**
- **Basic explanation of the analysis result**
- **User registration and login**
- **Text input for news articles and social media posts**
- **Credibility score display**

Technologies

- **Tools: Jupyter Notebook**
- **Technology: Python language**
- **Data Pre-Processing: Pandas, Numpy, NLTK, Doc2Vec**
- **ML Models: Naive Byes, SVM, Decision Tree, XGBoost, KNN, Logistic regression, Ensemble models, TensorFlow/Keras, Neural Network**

USER STORIES

“As a concerned citizen, I want to quickly verify the authenticity of the news articles I come across, so that I can avoid sharing misinformation with my friends and family”

“As a journalist, I want to see a detailed breakdown of why content was classified as fake or real.”

“As a student, I want to verify the information I use in my research papers, so that I can confidently cite reliable sources and improve the quality of my academic work.”

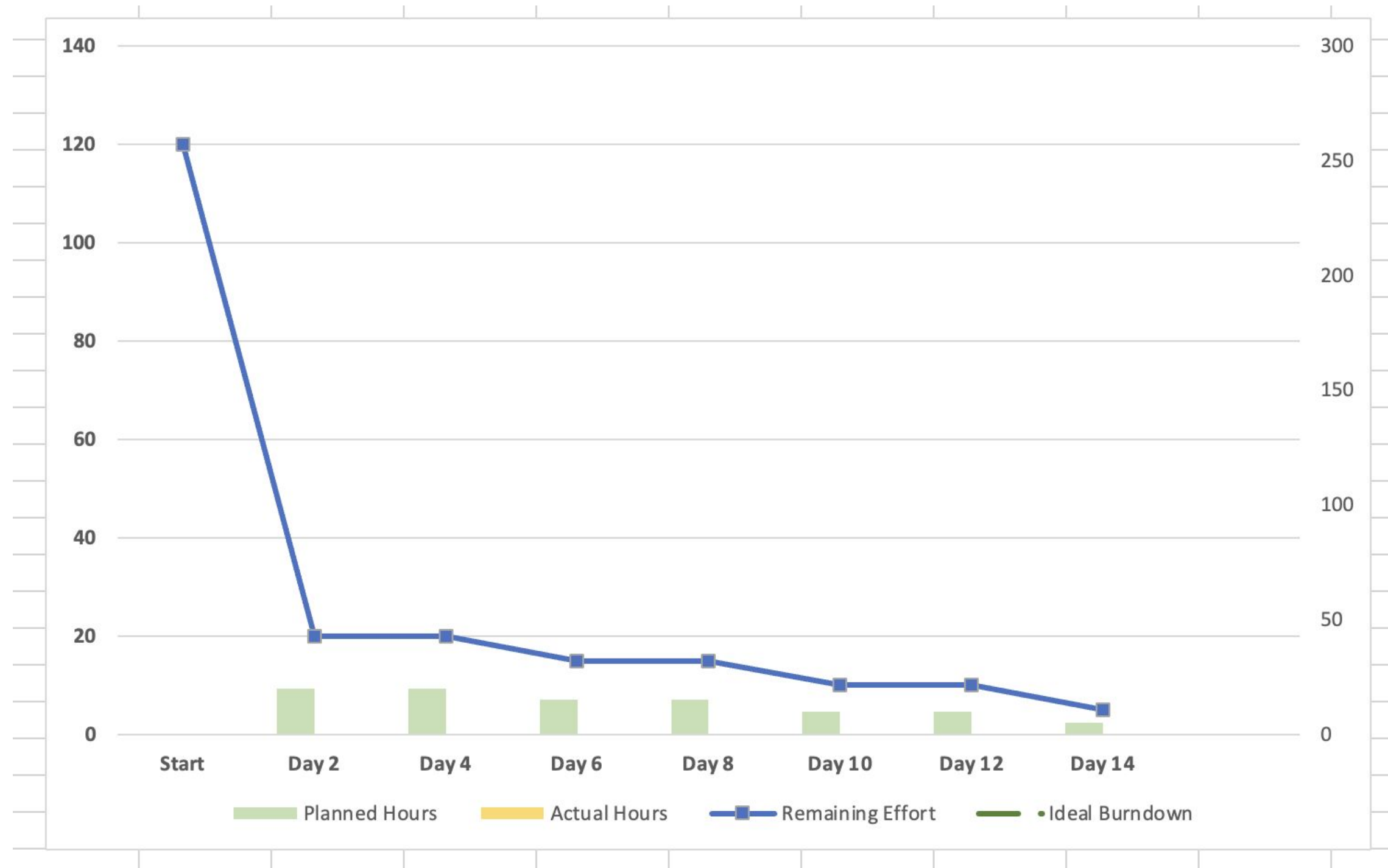
PROJECT BACKLOG

		Features/Task	Story Point	Status
	1	Basic web interface	5	Pending
	2	User authentication	3	Pending
	3	Analysis report	4	Pending
	4	URL input support	6	Pending

TEST CASES

	Test	Category	Pass Criteria	Status
1	Test with various text lengths	Input Validation	System should handle all text lengths without errors	Pending
2	Test with special characters and different languages	Input Validation	System should correctly process and analyze input regardless of language or characters	Pending
3	Test user registration with valid inputs	Authentication	Registration should succeed with valid inputs	Pending
4	Test user registration with invalid inputs	Authentication	Registration should fail with invalid inputs	Pending
5	Test login with correct credentials	Authentication	User should successfully log in with correct credentials	Pending
6	Test login with incorrect credentials	Authentication	User should not be able to log in with incorrect credentials	Pending
7	Test with known fake news dataset	Analysis Accuracy	System should correctly identify fake news	Pending
8	Test with known real news dataset	Analysis Accuracy	System should correctly identify real news	Pending
9	Test response time for analysis	Performance	System should analyze input and return results within an acceptable time frame	Pending
10	Test concurrent user handling	Performance	System should handle multiple users simultaneously without performance degradation	Pending

SPRINT BURNDOWN CHART



RETROSPECTIVE

What Went Well:

- **Good Teamwork, Taking iterative approach allowed for flexibility and adaptation to changing project needs.**
- **Successful implementation of machine learning models.**

Areas for Improvement:

- **Enhance model accuracy and reduce false positives.**
- **Optimize processing speed for real-time analysis**

Lessons Learned:

- **Regular communication is key to maintaining project momentum.**
- **Early integration testing helps identify and resolve issues faster**

Thank
you

