Group Project
Unit: Advance Natural Language Processing
Unit Code: COMP8420

Submitted by:
Abhishek Gupta (47745649)
Samyak Jain (47688408)

Introduction

Welcome to the AutoClaim, an innovative solution designed to revolutionize the way insurance claims, particularly car accident claims, are processed and analyzed. This report is a showcase of integrating cutting-edge technologies such as Anthropic's AI models, Streamlit for user interfaces, and PyMuPDF for PDF handling, AutoClaim provides a seamless and efficient way to analyze car accident data, extract insights, and generate comprehensive reports.

The insurance industry has long relied on manual processes for handling car accident claims, which can be time-consuming, error-prone, and inefficient. With the advent of artificial intelligence and machine learning, there is a significant opportunity to automate and optimize these processes. AutoClaim aims to address this challenge by leveraging cutting-edge AI technologies to analyze car accident reports, repair cost estimates, and insurance claim documents accurately and efficiently.

The primary objectives of the AutoClaim project are as follows:
- Automate the analysis of car accident reports, repair cost estimates, and insurance claim documents.
- Improve the accuracy and efficiency of the claim analysis process.
- Provide a user-friendly web application for insurance claim officers to upload and analyze relevant documents.
- Generate comprehensive reports that support data-driven decision-making in the claims process.
- Demonstrate the potential of AI in transforming the insurance industry.

Link to the project:
https://github.com/47688408/GroupProject_COMP8420

Methodology

About Dataset

AutoClaim utilizes a custom (specifically created for a particular purpose) dataset consisting of 11 images of car accidents, repair cost estimates, and insurance claim documents. This dataset has been carefully curated to provide a representative sample of real world documents for testing the system.

Technologies Implemented:

A. **Anthropic AI Models:** At the core of AutoClaim's analysis capabilities lies the Anthropic API and its powerful claude-3-opus-20240229 model. This state-of-the-art AI model is capable of understanding and analyzing complex text and image data, making it well-suited for tasks such as:

- Text Analysis: Extracting relevant information from insurance claim documents, repair estimates, and other textual data sources.
- Image Analysis: Analyzing car accident images to describe the visible damage and provide detailed accident reports.
- Data Synthesis: Combining and integrating insights from multiple data sources to generate comprehensive accident analysis reports.

The claude-3-opus-20240229 model leverages advanced natural language processing (NLP) and computer vision (CV) techniques to deliver accurate and insightful analysis, enabling informed decision-making in the claim processing workflow.

B. **Streamlit:** Streamlit is a powerful Python library used for building interactive web applications. In AutoClaim, Streamlit is utilized to create a user-friendly interface that allows users to upload car accident images, repair estimates (in PDF or image format), and insurance claim documents (in PDF or image format).

C. **PyMuPDF:** PyMuPDF is a robust Python library for handling and processing PDF documents. In AutoClaim, PyMuPDF plays a crucial role in extracting text from PDF files, displaying PDF pages within the Streamlit application, and processing PDF data efficiently.

Evaluation method

We did not have as such ground truth or correct output from which we could evaluate metrics such as accuracy, recall, etc. Instead we evaluated the response for different prompts by another LLM to identify which works out best. Also, there was no model training involved.

This evaluation method led us to different but interesting findings such as which and why specific prompts were better than others. Figuring out which prompts works best saves a lot of resources for further experimentation.

Pairwise comparison for best summarization and analysis

Generate Outputs:

- Have different prompts (5 in our case)
- Generate reports from different prompts for a given task (e.g., summarizing car accidents)

Evaluation:

- Employ another language model (Perplexity in our case) to act as the evaluator for larger scale assessments.
- This allows for evaluating many summaries or prompts efficiently.

The basis for comparing the reports are:

A. Comprehensiveness of Analysis: How thorough and detailed the summary is.
B. Objectivity and Impartiality: The neutrality and fairness of the summary.
C. Evidence-based Assessment: The extent to which the summary relies on factual data.
D. Clarity and Structure: How clear and well-organized the summary is.
E. Actionable Recommendations: Practical suggestions or next steps provided.
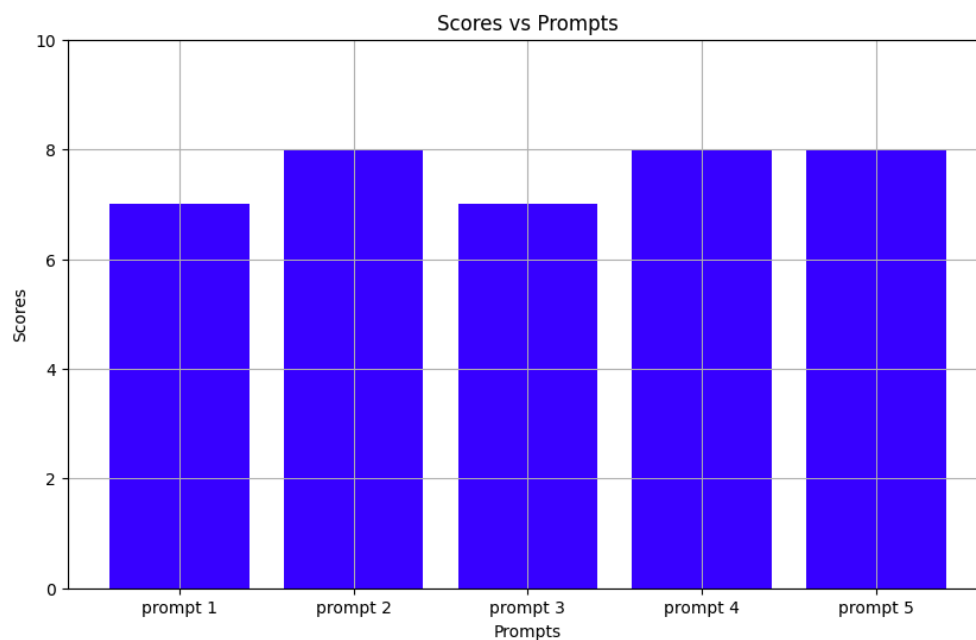

Fig1: Prompts score comparison

Technical Implementation

AutoClaim's technical implementation relies on Python and various open-source tools. Here's a streamlined breakdown of its key components:

A. Web Interface: Built with Streamlit, allowing users to upload documents and images directly from Python scripts. The **main.py** file powers this interface, enabling file uploads, image display, and report generation.

B. Image Analysis: Utilizes the Anthropic API via the **analyse_car_accident_images.py** file. Images are encoded into base64 format using **image_encoder.py**, then analyzed by the Claude AI model to describe car accident scenes.

C. Accident Analysis: The final report (PDF) file combines results from image, repair estimate, and insurance claim analysis. It instructs the Claude AI model to generate a comprehensive report, including accident analysis, repair cost estimates, and insurance claim recommendations.

D. Report Generation: The create_pdf function in **main.py** uses the FPDF library to generate a PDF report with formatted text output from the accident analysis.
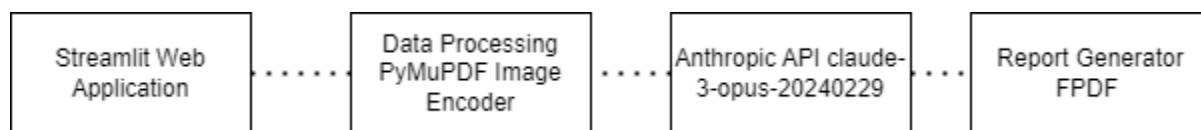
Effectiveness of AutoClaim

Manual report generation by insurance agents is a time-consuming and labor-intensive process, often leading to higher costs and potential errors. According to a study by McKinsey & Company, manual data entry and report generation can account for up to 30% of an insurance agent's workload (Balasubramanian et al., 2021). Additionally, a report by Accenture suggests that manual processes in the insurance industry can lead to operational inefficiencies and increased costs of up to 25% (Accenture Strategy & Consulting, 2021). Furthermore, manual report generation is prone to human errors, inconsistencies, and potential non-compliance with industry regulations.

In contrast, the AutoClaim system generated the report in a span of 10-12 seconds, which represents a significant time saving compared to manual processes. By automating report generation and reducing the time required from hours to just 10-12 seconds, the AutoClaim system can potentially lead to substantial cost savings by improving operational efficiency

and reducing labor costs. The AutoClaim system, by generating reports in a matter of seconds, can ensure consistency, accuracy, and compliance with relevant regulations, thereby improving the overall effectiveness of the report generation process.

Contribution

The dataset creation task was evenly split between the team members, with each taking on 50% of the task. Following the collection:

Abhishek worked on the image encoding function to encode images which was crucial to send to Claude APi message requests. Abhishek also worked on evaluation methods (pairwise comparison) to assess which report has best analysis. For pairwise comparison, the response(reports) generated by Claude was fed into Perplexity.ai to compare the results on the specific set of criteria and project's effectiveness in terms of manual report generation by insurance agent and automatic report generation by AutoClaim regarding cost, time and consistency.

Samyak worked on optimizing python code for reusability, building Claude APi message requests, selecting which pdf library among Pymupdf, PyPDF and FPDF is best suited to generate comprehensive reports, the selection of different prompts and building web application interface with Streamlit.

Both members collaborated closely on the development of **main.py** and jointly contributed to the final project report.

Future Enhancements

AutoClaim provides significant advantages in automating car accident claim processing, with potential future enhancements such as automated claim decision-making, advanced analytics, scalability, multi-language support, mobile applications, and IoT integration. These improvements can further streamline operations, enhance user experience, and broaden market reach. AutoClaim offers benefits like increased efficiency, accuracy, cost savings, transparency, scalability, and consistency by reducing manual intervention, leveraging AI, and ensuring fair and reliable claim assessments. Additionally, PyMuPDF facilitates efficient text extraction and analysis of PDF documents within AutoClaim, enhancing data processing and report generation capabilities.

Conclusion

AutoClaim revolutionizes insurance with AI, computer vision, and NLP, streamlining car accident claim analysis for efficiency and accuracy. Integrating cutting-edge tech like Anthropic's Claude AI, it processes diverse data sources comprehensively. This ensures no critical information is missed, optimizing claim assessments. As technology advances, AutoClaim becomes pivotal in reshaping insurance, benefiting companies and customers alike. Insurers can cut costs, boost satisfaction, and maintain competitiveness. Customers enjoy faster processing, transparent decisions, and fair settlements. Its scalable architecture and machine learning adaptability make AutoClaim future-proof. Continuous improvement fine-tunes models for even better efficiency and accuracy over time.

In summary, AutoClaim showcases the power of AI and advanced tech in enhancing business processes, driving innovation, and delivering value to all stakeholders.

References

Balasubramanian, R., Libarikian, A., & McElhaney, D. (2021, March 12). Insurance 2030—The impact of AI on the future of insurance. McKinsey & Company. https://www.mckinsey.com/industries/financial-services/our-insights/insurance-2030-the-impact-of-ai-on-the-future-of-insurance

Accenture Strategy & Consulting. (2021). Accelerating automation for insurance. https://www.accenture.com/content/dam/accenture/final/a-com-migration/pdf/pdf-148/Accenture-Accelerating-Intelligent-Automation-Insurance.pdf

Appendix

Web interface :

Upload Images

Upload Car Accident Image

☁ **Drag and drop file here**
Limit 200MB per file • JPG, JPEG, PNG

Browse files

📄 **7.jpg** 375.1KB ✕

Car Accident Image

Upload Car Repair Estimate Document

☁ **Drag and drop file here**
Limit 200MB per file • JPG, JPEG, PNG

Browse files

📄 **7.jpg** 65.0KB ✕

```
Car Part, Estimated Cost (USD)

Front Bumper: $600 USD
Hood: $800 USD
Front Grille: $350 USD
Headlights: $500 USD (both)
Front Fenders: $700 USD (both)
Windshield: $400 USD
Airbags (Driver and Passenger): $1000 USD
Front Suspension: $1200 USD
Wheels and Tires (Front): $800 USD
Radiator and Cooling System: $600 USD
Engine Repairs: $2500 USD

Total Estimated Repair Cost: $9450 USD
```

Upload Car Insurance Claim Document

Drag and drop file here
Limit 200MB per file • JPG, JPEG, PNG

Browse files

7.jpg 198.6KB ✕

```
Insurance Claim Document:

Policyholder Information:
Name: Sarah Thompson
Address: 456 Oak St, Anytown, AN 54321
Contact Information: (555) 987-6543 / sarahthompson@email.com
Policy Number: POL-984251

Vehicle Information:
Make: Mercedes-Benz
Model: C-Class
Year: 2015
VIN: 7892143
License Plate Number: 7521ABC

Incident Information:
Date and Time of the Accident: 2024-03-12 09:35:21
Location of the Accident: Highway 101, near Exit 42
Brief Description of the Accident: Collision with a guardrail on the right side of the highway.
Weather and Road Conditions: Rainy, wet road
Police Report Number: REP-35671
Other Involved Parties or Witnesses: None

Damage Information and Repair Estimate:

- Front Bumper: $600 USD
- Hood: $800 USD
- Front Grille: $350 USD
- Headlights: $500 USD (both)
- Front Fenders: $700 USD (both)
- Windshield: $400 USD
- Airbags (Driver and Passenger): $1000 USD
- Front Suspension: $1200 USD
- Wheels and Tires (Front): $800 USD
- Radiator and Cooling System: $600 USD
- Engine Repairs: $2500 USD

Total Estimated Repair Cost: $9450 USD

Declaration:
I hereby declare that the information provided above is accurate to the best of my knowledge and belief.

Signature: _____

Date: [Date]
```

Car Insurance Claim

Generate Report

Output :

Accident Analysis Results:
Car Accident Image Analysis Result:
Based on the image, the red Mazda CX-5 SUV has sustained significant front-end damage in what appears to be a collision. Here is a detailed accident report describing the visible damage:

Front Bumper: The front bumper is severely damaged and partially detached from the vehicle. It is crumpled inward and misshapen from the impact.

Hood: The hood is crumpled and bent upwards, likely from the force of the collision. It no longer aligns properly with the front of the vehicle.

Grille and Front Fascia: The front grille, logo and surrounding front fascia components are badly damaged and broken. Pieces of the grille and trim are missing or hanging off.

Headlights: Both front headlight assemblies have shattered and are non-functional due to the extensive damage to the front end.

Fenders: The front fenders on both the driver and passenger side are crumpled and deformed, especially around the wheel well openings.

Windshield: Although not shattered, the windshield has several cracks likely from the force and twisting of the vehicle's frame and body panels during the impact.

Airbags: The front driver and passenger airbags appear to have deployed, indicating the sensors detected a severe frontal collision.

In summary, this Mazda CX-5 has suffered major front end damage rendering it undriveable. The vehicle will require extensive body work, mechanical repairs and replacement of most front components before being roadworthy again, if repairable at all. Injuries to occupants are likely given the severity of the impact. A full investigation into the circumstances and causes of the accident is recommended.
Repair Estimate Analysis Result:
The car repair estimate breaks down the costs for various parts of the vehicle:

- Front bumper: $800
- Hood: $600
- Grille: $250
- Headlights (both): $400
- Radiator: $700
- Engine repair/replacement: $2,500
- Front fenders (both): $600
- Windshield: $400

The total estimated repair cost sums up to $6250 USD.

Key takeaways:
- Engine repair/replacement is the most expensive item at $2,500
- Body parts like the bumper, hood, and fenders also contribute significantly to the total
- Headlights and windshield replacement add $800 combined
- The front end of the vehicle seems to have sustained the most damage based on the parts listed

Overall, this appears to be an extensive and costly repair, likely from a major front-end collision. The estimate thoroughly itemizes the various components needing repair or replacement.
Insurance Claim Analysis Result:

This insurance claim form provides details about an auto accident involving a 2018 Mazda CX-5 on June 8, 2024 at the intersection of Main St. and Oak Ave. in Townville, CA. The incident resulted in a front-end collision causing significant damage to the vehicle's front end.

Weather and road conditions were clear skies and dry at the time. The police report number for the incident is TPD-2024-06081122.

The form breaks down the estimated repair costs, totaling $6250. This includes $800 for the front bumper, $600 for the hood, $250 for the grille, $400 for headlights, $700 for the radiator, $2,500 for engine repair/replacement, $600 for front fenders, and $400 for the windshield.

The policyholder, John Smith, must sign and date the form to declare the provided information is accurate to the best of their knowledge.