esprit Se former autrement HONORIS UNITED UNIVERSITIES	TUNISIAN REPUBLIC ***** MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH ***** GENERAL DIRECTORATE OF TECHNOLOGICAL STUDIES ***** International Data Science Cclass	

Project REPORT

Presented By:

Ben Ammar Mohamed Amine

Bennour Ines

Bouguecha Feriel

Chaabane Ela

Marzouki Mohamed Amine

Soltani Ahmed

Subject

DueDiligence

Supervised by:

Mr Bilel Farah

Mrs Jihene Jebri

Project Provider: Value

Academic year: 2024/2025

1. Introduction

In the ever-evolving world of digital finance, conducting thorough due diligence is essential to ensure the legitimacy, security, and long-term viability of cryptocurrency projects. As blockchain technology continues to disrupt traditional financial models, investors, regulators, and industry participants must adopt robust evaluation frameworks to mitigate risks and seize opportunities effectively.

This report, leveraging the principles of the **Team Data Science Process (TDSP)**, delves into the initial phases of a **comprehensive due diligence framework** designed for assessing cryptocurrency projects. The objective is to create a structured, data-driven approach that enhances transparency, security, and compliance within the crypto ecosystem.

The project focuses on the development and implementation of an **automated due diligence system**, utilizing cutting-edge technologies such as **AI-powered risk assessment, blockchain analytics, and regulatory compliance automation**. As the cryptocurrency space matures, traditional evaluation methods are giving way to more sophisticated **machine learning and NLP-based solutions**, ensuring that investors and stakeholders can make informed decisions with greater accuracy and efficiency.

This report navigates through the **initial phase of the TDSP**, offering a detailed exploration of the **crypto industry landscape, risk assessment methodologies, and business understanding**. By adhering to a structured and iterative approach, this framework aligns with **industry best practices, regulatory standards, and technological advancements**, setting a new benchmark for due diligence in the digital asset space.

2. State of art

Created in 2019 ,Value is a consulting firm in strategy and digital transformation, committed to creating sustainable value for society and the economy alongside its partners and customers.
Value is a true incubator of skills, helping its employees to develop and reach their full potential by contributing to major projects with a very high impact while developing their business and technological expertise within its various division.

Services :
Digital Factory: Designs and builds digital solutions to streamline business processes and enhance customer experiences.
Data Factory: Collects, processes, and manages data pipelines to deliver actionable insights.
Advanced Analytics & AI Services: Leverages advanced analytics and AI technologies to drive data-driven decisions and innovation.
Strategic Advisory: Provides expert guidance to help businesses align strategies with market trends and achieve long-term goals.

2.1. Actors

The cryptocurrency ecosystem comprises a diverse range of actors including:

- **Investors:** Require trustworthy insights to make confident decisions on cryptocurrency and digital asset investments.
- **Venture Capital Firms:** Need data-driven evaluation frameworks to identify promising blockchain startups and digital asset projects.

2.2 Competition

Competitor	Core Services	Strengths	Weaknesses
DueDil	Company information services	AI-driven insights	Limited focus on crypto funds
AutoGen AI	Automated report generation	Simplifies reporting	Basic compliance features
Chainalysis	Blockchain analytics	Advanced risk monitoring	Expensive and lacks customizable reporting
Elliptic	Risk monitoring for crypto	Strong fraud detection	Limited report generation capabilities

2.3 Swot Analysis

A SWOT analysis of the current crypto due diligence landscape is outlined below:

Strengths:

- Advanced analytics and AI-driven models provide rapid and accurate insights.
- Increased transparency in blockchain transactions fosters trust among stakeholders.
- Growing adoption of digital assets offers a wealth of data for analysis.

Weaknesses:

- Regulatory uncertainty can hinder widespread adoption.
- Rapid technological changes may outpace current analytical frameworks.
- High complexity and volatility of crypto markets can challenge model accuracy.

Opportunities:

- Expansion into emerging markets with growing interest in cryptocurrencies.
- Integration of cutting-edge AI and data science methodologies.
- Potential for strategic partnerships with financial institutions and tech firms.

Threats:

- Increasing regulatory scrutiny and potential legal challenges.
- Competition from both established financial entities and innovative startups.
- Cybersecurity risks and potential market manipulation.

Our Proposed Solution:

1. **Data Automation:** Streamlined data collection from diverse sources.
2. **Risk Detection:** Advanced AI models for early fraud detection.
3. **Compliance Checks:** Real-time regulatory analysis.
4. **Actionable Reports:** Customizable, insightful reports tailored to investors and VCs.
5. **Investment Insights:** Predictive analytics for identifying high-potential opportunities.

2.4 Overview of TDSP Methodology

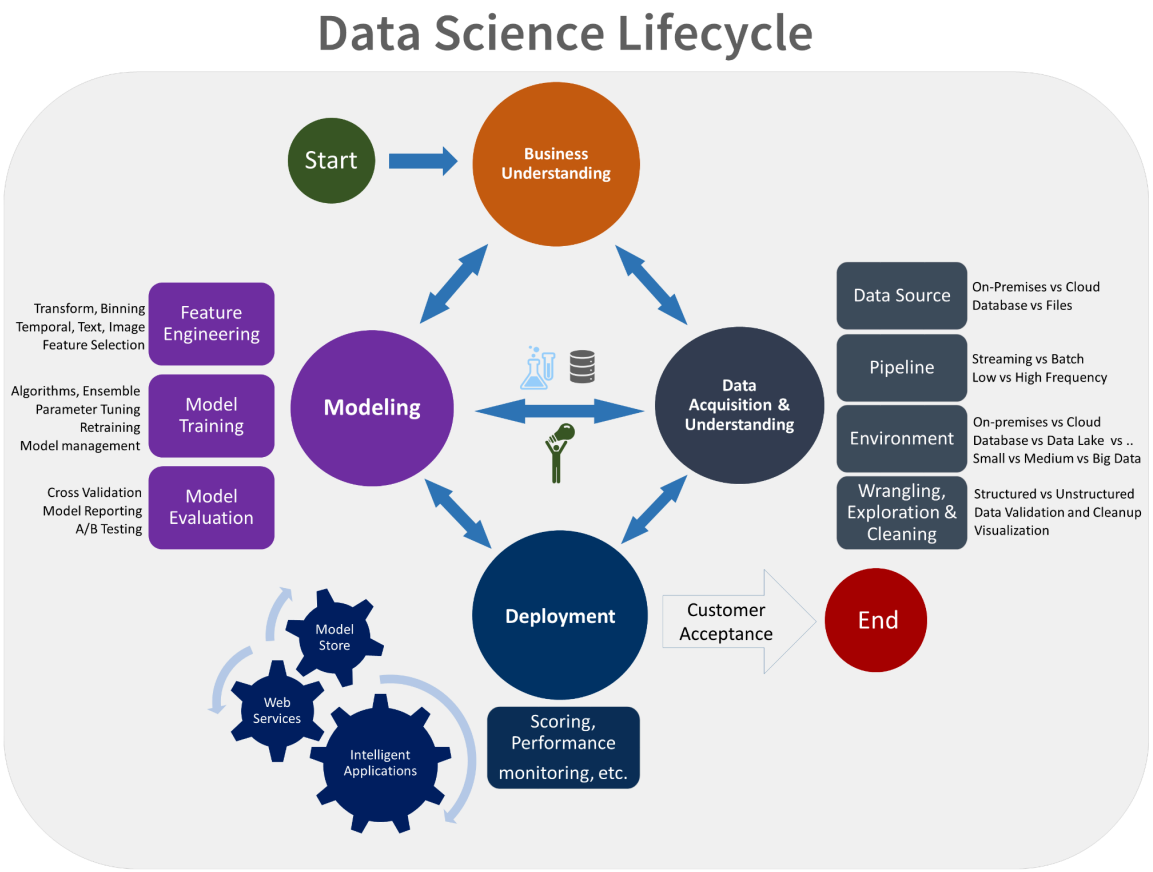

image 1: TSDP Life Cycle

- **Methodological Approach:**
  The project follows the Team Data Science Process (TDSP), an agile and iterative methodology that ensures alignment between business objectives and technical execution.
- **Phases Covered:**
  - **Business Understanding:** Define the problem, goals, and stakeholder requirements.
  - **Data Acquisition & Understanding:** Gather and preprocess relevant crypto data.
  - **Modeling:** Develop and refine models to support due diligence.
  - **Deployment:** Implement the solution in a production environment, ensuring scalability.

    - **Customer Acceptance:** Engage with stakeholders for feedback and validate the solution against real-world scenarios.
- **Benefits:**
  TDSP's structured approach facilitates collaboration, continuous improvement, and effective risk management throughout the project lifecycle.

3. Business Understanding

3.1. Business Domain

This project operates at the crossroads of **digital finance, blockchain, and data analytics**. It aims to enhance **due diligence** processes in the **cryptocurrency sector** by using **advanced analytics** to evaluate the credibility of crypto projects. Key focus areas include:

- Assessing **project viability** based on historical trends, market activity, and technical robustness.
- Identifying **security risks**, such as vulnerabilities in smart contracts or historical fraud patterns.
- Ensuring **regulatory compliance** by integrating automated checks against global financial regulations.

Due Diligence in Finance and Its General Application

Definition:

Due diligence in finance is the careful and thorough process of investigating and verifying all relevant information about an investment or business opportunity before making any significant decisions. It involves evaluating financial data, legal compliance, operational aspects, and potential risks to ensure that all material facts are accurately understood.

General Application:

The principles of due diligence extend beyond traditional finance and can be applied to various projects, including those involving data collection and analysis. In any project, whether it involves digital assets, operational improvements, or new technology implementations, due diligence helps to:

- **Ensure Data Integrity:** Verify that the data sources are reliable and the information is accurate.
- **Identify Risks:** Detect potential issues early on by carefully reviewing and analyzing all available information.
- **Support Decision-Making:** Provide a solid foundation for informed decisions by summarizing verified insights in clear, actionable reports.
- **Enhance Transparency:** Maintain a clear record of the evaluation process to build trust and accountability.

By leveraging **data-driven techniques**, the project helps investors and stakeholders make **informed and risk-mitigated decisions** in the crypto space.

3.2. Business Problem

The primary business challenge addressed by this project is the inherent risk and uncertainty in the rapidly evolving crypto market. Traditional due diligence methods often fail to capture the dynamic nature of digital assets, leading to potential investment risks and regulatory non-compliance. There is a critical need for a structured, data-driven approach to assess the multifaceted risks associated with crypto projects.

3.3. Objectives

1. Automate and Optimize Due Diligence

- **Business Objective:** Reduce manual effort by **50%** while maintaining high-quality insights.

Due diligence in the crypto space often involves time-consuming manual research, including reading whitepapers, analyzing project roadmaps, and assessing tokenomics. By automating repetitive tasks such as data collection, document analysis, and sentiment tracking, we can significantly reduce workload while improving efficiency.

- **Data Science Solution:** Develop **automated data pipelines** that efficiently process large-scale cryptocurrency datasets from multiple sources, ensuring seamless integration and rapid analysis.

2. Provide Real-Time, Actionable Insights

- **Business Objective:** Deliver insights with **90% accuracy**, enabling data-driven decisions.

Investors and analysts require timely insights to make informed decisions in the fast-moving crypto market. Static reports become outdated quickly, leading to missed opportunities or exposure to high-risk investments.

- **Data Science Solution:** Fine-tune **Large Language Models (LLMs)** using **crypto-specific datasets**, improving the accuracy of market trend analysis, project assessments, and fraud detection. Real-time data processing and dynamic risk scoring will ensure that stakeholders receive the most current and relevant information.

3. Standardize Evaluations Across Crypto Funds

- **Business Objective:** Ensure **100% consistency** in fund assessments for better decision-making.

Current evaluation methods vary widely among analysts, leading to inconsistencies in fund ratings and due diligence conclusions. A standardized framework is required to ensure reliability and comparability across different funds.

- **Data Science Solution:** Use **LLMs to generate structured evaluation frameworks** and refine prompting strategies to eliminate subjective biases. This approach will ensure that each fund is assessed based on uniform criteria, improving comparability and transparency.

4. Mitigate Investment Risks

- **Business Objective:** Identify **95% of potential fraud cases** to reduce financial exposure.

The crypto industry is rife with fraud schemes, including rug pulls, wash trading, and Ponzi-like structures. Identifying these risks early can prevent significant financial losses for investors.

- **Data Science Solution:** Apply **Natural Language Processing (NLP)** to analyze project documentation, smart contracts, and on-chain transactions for red flags. Additionally, implement **anomaly detection models** to flag unusual patterns in fund behavior, significantly improving fraud detection accuracy.

5. Ensure Full Regulatory Compliance

- **Business Objective:** Achieve **100% automated compliance checks** on funds.

  Crypto projects often struggle to comply with evolving regulatory requirements, leading to legal risks and potential sanctions. Manual compliance verification is slow and prone to errors.

- **Data Science Solution:** Integrate **regulatory APIs** (e.g., **SEC EDGAR, FINRA, FATF databases**) to automate compliance verification. This ensures that all crypto funds adhere to the latest financial regulations and standards, reducing regulatory risks.

6. Deliver Interactive and Bespoke Reports

- **Business Objective:** Generate **high-quality reports** with **85% user satisfaction**, enhancing trust.

Stakeholders need clear, insightful, and well-structured reports to understand risk assessments and investment recommendations. Manual report generation can be inconsistent and time-consuming.

- **Data Science Solution:** Automate the creation of **PPTX reports** with **data-backed insights** and visualizations, ensuring professional, engaging, and easily interpretable presentations. Interactive features such as drill-down capabilities and real-time updates will further improve user experience and satisfaction.

By implementing these objectives, the project aims to transform crypto due diligence into a **scalable, efficient, and highly accurate process**, ensuring **data-driven investment decisions and regulatory compliance**.

3.4. Requirements

3.4.1. Functional Requirements

- Collect and aggregate data from multiple sources, including APIs, web scraping, and regulatory databases.
- Process and analyze large volumes of blockchain data in real-time.
- Generate comprehensive reports that include risk assessments, compliance checks, and market analysis.
- Provide an interactive dashboard for stakeholders to monitor key performance indicators.

3.4.2. Non Functional Requirements

- Scalability: The system should handle increasing volumes of data as the market grows.
- Performance: Real-time data processing and rapid report generation are essential.
- Security: Robust measures must be in place to protect sensitive financial and personal data.
- Reliability: The system should maintain high availability and accuracy under varying market conditions.

3.5. SDGs

The project aligns with several Sustainable Development Goals (SDGs), including:

- SDG 8 (Decent Work and Economic Growth): By promoting transparent and secure financial practices.
- SDG 9 (Industry, Innovation, and Infrastructure): Through the integration of cutting-edge data analytics and blockchain technologies.
- SDG 16 (Peace, Justice, and Strong Institutions): By enhancing regulatory compliance and fostering trust in digital finance.

4. Relative Requirements

4.1. Business and Data Science Objectives' Metrics

- Risk Assessment: The development of predictive models to evaluate the security and compliance risks of crypto projects.
- Market Competitiveness: The ability to identify and analyze emerging market trends and investor sentiment.

- Regulatory Compliance: The automated integration of regulatory standards and real-time compliance checks.
- Transparency: The creation of explainable, transparent analytical models that foster stakeholder trust.
- Efficiency: The development of an automated due diligence system that significantly reduces manual effort.
- Investment Performance: The implementation of forecasting models to estimate future returns and project viability.
- Stakeholder Trust: The application of fairness-aware machine learning techniques to ensure unbiased assessments.

4.2. Requirements' Metrics

Functional Metrics:

- Application of the crypto due diligence model for comprehensive risk and compliance assessment.
- Categorization of crypto projects based on risk segments and market potential.
- Maintenance of detailed audit trails for compliance and accountability.
- Integration with diverse external data sources (APIs, web scraping, regulatory feeds) to enhance accuracy.

Non-Functional Metrics:

- Adaptability to changes in market dynamics, regulatory frameworks, and data sources.
- Scalability to handle varying volumes of blockchain and market data efficiently.
- Reliability with a high target availability to ensure continuous service.
- Security adherence to industry standards for robust data protection.
- Model explainability to ensure transparency in decision-making processes.

4.3. Models' Evaluation Metrics

For evaluating the performance of AI and machine learning models, the following metrics will be used:

1. Data Collection and Engineering Pipeline Metrics

- **Data Completeness:** Percentage of missing or incomplete data points.
- **Data Freshness:** Recency of the collected data.
- **Processing Efficiency:** Time taken per data batch.
- **Error Rate:** Number of failed operations during pipeline execution.

2. Generative Question Bank Metrics

- **Relevance Score:** Alignment with due diligence goals, measured through semantic similarity tools.
- **Diversity Metrics:** Use **Self-BLEU** or **Distinct-n** to measure question uniqueness.
- **Difficulty Scoring:** Combination of readability metrics (like Flesch-Kincaid) and expert ranking.

3. Q&A System Metrics

- **Accuracy:** F1 Score and Exact Match (EM) with a benchmark set of answers.
- **Perplexity:** Measures fluency and confidence of the model in generating responses.
- **Relevance:** Semantic similarity (cosine similarity) between generated and reference answers.
- **Response Time:** Latency for generating answers.
- **Safety:** Evaluation of hallucinations, bias, and harmful content.

4. Report Generation Module (PPTX) Metrics

- **Content Coverage:** Proportion of key due diligence aspects captured in the generated reports.
- **Readability:** Use readability indices to measure slide clarity.
- **Formatting Consistency:** Evaluate slide layout coherence and design uniformity.
- **Human Evaluation:** Domain experts assess utility and completeness.

5. Conclusion

In conclusion, the proposed due diligence framework represents a significant advancement in the evaluation of cryptocurrency projects. By integrating data-driven methodologies with automated analytics, the system not only enhances risk assessment but also aligns with industry best practices and regulatory standards. This comprehensive approach is designed to provide stakeholders with transparent, actionable insights, paving the way for more secure and efficient investments in the dynamic world of digital finance.

Future work will focus on refining analytical models, expanding data integration capabilities, and continuously adapting the system to meet emerging challenges in the rapidly evolving crypto ecosystem.