**Introduction**

Welcome to our presentation on leveraging People Analytics for enhancing HR decision-making. In today's data-driven environment, traditional tools like Excel no longer suffice due to their limitations in scalability and efficiency. Our project utilizes predictive analytics and machine learning to address these gaps. We aim to develop sophisticated models that can predict voluntary employee churn and optimize hiring processes effectively, using a dataset encompassing 25,856 employee records. This strategic initiative is designed to provide leadership with actionable insights to refine HR operations and talent management.

**Problem Statement**

In facing the challenges of HR analytics, our project specifically targets enhancing decision-making capabilities by integrating advanced analytical methods. We manage a wide array of data, ranging from basic employment details to comprehensive workforce demographics and compensation frameworks, collected over several years. The primary goal is to leverage this rich dataset through predictive analytics to not only forecast hiring needs with high accuracy but also to identify potential voluntary turnovers. This approach is aimed at crafting a robust and sustainable talent management strategy that aligns with our overall business objectives.

**Data Cleaning and Preparation Process**

Our data cleaning methodology has been exhaustive and precise. Starting with data profiling, we meticulously identified and recorded any quality concerns. We then tackled missing data using a blend of statistical methods and domain expertise to ensure completeness. To guarantee the dataset's uniqueness, we eliminated duplicate records and applied normalization techniques to prepare the data for sophisticated analytics. The process also involved feature engineering to enhance our dataset, particularly focusing on incorporation of financial metrics that are vital to our analysis.

A screenshot of a data report

Description automatically generated

The visualization illustrates the thoroughness of our data cleaning process, underlining both the deletion and imputation of records due to issues such as missing values and outliers. Initially comprising 25,995 records, the dataset was distilled to 25,856 records post-cleaning, with 139 records, or 0.53%, removed. Additionally, 1,448 entries were imputed to address null values in 'Compa Ratio' and 'Base Pay', accounting for 5.57% in each column. A logarithmic comparison of outliers presents a clear before and after view, showing a notable reduction in 'Compa Ratio' outliers from 6.73% to 1.26%, demonstrating an improved data quality essential for reliable analytics.

**Data Splitting and Model Preparation**

**A diagram of a company

Description automatically generated with medium confidence**

Our HR dataset of 25,856 records was carefully segmented into two distinct groups: 14,168 records of terminated employees and 11,688 records of active employees, delineated for historical and predictive analysis. A systematic approach was adopted for model training and validation: the terminated employee dataset was split into an 80/20 partition for training and validation respectively, while the active employee data was set aside as a holdout set to assess the model’s effectiveness in real-world conditions. Feature engineering was meticulously carried out to design features that capture the predictors of employee turnover, with an emphasis on job-related, demographic, and financial factors. This rigorous setup aimed at refining our model to focus on voluntary terminations, ensuring unbiased predictive capabilities by excluding fields like 'Termination Date', 'Termination Type', and 'Termination Reason'.

**Model Overview and Performance Comparison**

In our analysis, we evaluated four advanced models: Random Forest, XGBoost, AdaBoost, and CatBoost, each contributing unique strengths to our predictive analytics capability. For instance, the Random Forest model, with its ensemble of decision trees, provided stable predictions with an accuracy of 76%. In contrast, XGBoost stood out for its handling of complex data interactions, achieving an impressive accuracy of 78% and a high ROC-AUC score of 86.7%. A comprehensive comparison reveals CatBoost's superior performance in managing categorical data, boasting a high recall and precision and an accuracy of approximately 77.66%. This positions CatBoost favourably among the models, all of which displayed a tight performance range, indicating their robustness for HR analytics.

A screenshot of a graph

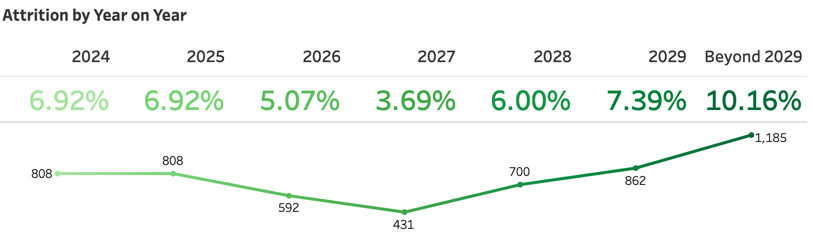
Description automatically generated

**Model Selection**

Considering the critical attributes for our HR analytics, CatBoost emerges as the optimal choice due to its excellent performance across several key areas. It achieves high recall, essential for accurately predicting voluntary terminations, and handles categorical features with greater efficiency, reducing the need for extensive pre-processing. Furthermore, CatBoost demonstrates resistance to overfitting, ensuring its predictions are reliable when applied to real-world scenarios. It offers operational efficiency with faster post-training prediction times, which is invaluable for time-sensitive HR decisions. The model's interpretability and transparent feature impact analysis adhere to legal and operational standards, making it an accountable choice. With its adaptability to various data distributions, CatBoost promises to remain a compatible and versatile tool for future HR predictive challenges.

**Predictive Analysis and Forecasting**

Our refined predictive analysis utilizes the CatBoost model to process both categorical and numerical features effectively. This model predicts termination probabilities, which we use to rank employees by predicted risk of voluntary departure. These insights are crucial for HR to implement targeted interventions aimed at reducing turnover, aligning with our strategic goal of enhancing employee retention through proactive measures.



**Note:** The Attrition rate for 2024, currently at 6.92%, reflects the exclusion of 120 employees i.e. 1.03% who have voluntarily terminated their employment during the current year.

A screenshot of a data presentation

Description automatically generated

The visual shows that of the total 11,668 active employees, a significant 46.16% voluntary turnover is predicted over the coming years, with an average tenure of 6.83 years per employee. The attrition rates are expected to be relatively stable in the immediate future, with a gradual increase to 10.16% by 2029 and beyond. This suggests a growing need for effective retention strategies as time progresses.

The attrition analysis breaks down further by gender and job category, indicating that males, who make up 67.68% of the workforce, and employees in Engineering and Operations roles are particularly prone to higher turnover rates. The geographic distribution of attrition highlights the top five countries affected, with the highest numbers in Thailand and the United States, pointing to regions where retention efforts might be most urgently needed. This detailed forecasting enables the organization to target retention initiatives more effectively, focusing on high-risk groups and regions to enhance overall employee retention and stability.

**Financial Strategy for Talent Acquisition**

The financial implications of not addressing voluntary churn are substantial. Our analysis projects a voluntary turnover of 1,616 employees over the next two years, leading to an estimated financial impact of approximately $66.8 million. To mitigate these impact, our strategic financial plan includes allocating a significant budget towards comprehensive hiring expenses. This includes funding for recruitment, training, and incentives such as relocation packages and joining bonuses, which are essential for maintaining competitive hiring practices.

**Lessons Learned**

Throughout this comprehensive project, we've encountered several enlightening lessons that emphasize the crucial role of detailed data preparation and the judicious selection of predictive models. We've learned that tuning the classification threshold, according to the balance between precision and recall, is paramount for actionable predictions-essential for strategic HR decisions. The deliberate segmentation of our data into training, validation, and test sets illuminated the importance of historical context, ensuring our models are robust and capable of generalizing well to predict future events. These steps have been instrumental in establishing a predictive analytics framework that supports proactive, data-informed HR strategies.

**3 Most Important Findings/Reasons Your Research Is Unique:**

1. **Advanced Predictive Modeling:** Our project goes beyond conventional analytics by integrating a suite of advanced predictive models. This approach allows us to tackle the intricacies of HR data, addressing complexities ranging from demographic variances to nuanced compensation structures.
2. **Strategic Analytical Translation:** The analytical outcomes we've obtained aren't merely numbers-they're insights. We've translated complex data patterns into strategic HR initiatives, paving the way for evidence-based decision-making that can refine talent acquisition, employee retention, and overall HR policy formulation.
3. **Proactive Predictive Strategy:** Our use of predictive analytics transcends traditional reporting. We've employed these insights proactively to pre-emptively address talent management and decision-making. This strategic foresight is what sets our HR operations apart, allowing us to not just respond to but anticipate and shape the future of our workforce dynamics.

**Conclusion**

In conclusion, our project exemplifies a comprehensive approach to managing the data lifecycle, from meticulous cleaning to strategic segmentation, leading to insightful model comparisons and precise, actionable predictions. The proactive strategies enabled by our predictive models have positioned HR to transition from a reactive to a proactive stance, effectively managing talent retention and facilitating strategic planning for future workforce needs.