**Enkhgegee Altangerel**

STATS 201

Problem Set 1

**Research Proposal**

***Predicting Career Outcomes: The Role of Internships and University Prestige Using Machine Learning and Social Network Analysis***

**1. Background and Motivation:**

Graduates' career trajectories are influenced by a complex interplay of factors, including educational background, internship experiences, and the prestige of their alma mater. Existing research indicates that graduates from prestigious universities often secure better employment opportunities and command higher salaries due to superior networking opportunities, institutional reputation, and educational quality (Jung & Kim, 2018; Drydakis, 2016). Similarly, internship experience has been shown to significantly enhance employability, with interns being more likely to receive job offers and higher starting salaries compared to non-interns (Gault, Leach, & Duey, 2010; NACE, 2017).

However, there is a need for deeper insight into how these factors interact, particularly in the context of **machine learning** and **social science**. Traditional statistical methods often struggle to capture the nuanced relationships between internships, university prestige, and career outcomes. Machine learning techniques, such as predictive modeling and social network analysis, offer powerful tools to uncover hidden patterns and provide actionable insights. For example, predictive models can identify the key factors that most influence career success, while social network analysis can reveal how professional networks mediate the impact of internships and university prestige.

This research aims to bridge this gap by leveraging **machine learning** to study the impact of internships and university prestige on career outcomes. Specifically, it seeks to answer the following questions:

* How can machine learning models predict the impact of internship experience on job offers and starting salaries for graduates from universities of varying prestige?
* What are the key factors (e.g., internships, university prestige, networking) that most influence career outcomes, and how can these factors be interpreted using explainable AI techniques?

By addressing these questions, this research contributes to the growing field of **machine learning for social science**, while providing practical insights for students, universities, employers, and policymakers. Additionally, it explores the ethical implications of using AI-driven tools for career counseling, ensuring that these technologies promote inclusivity and fairness.

**2. Research Question:**

This study seeks to answer the following core research questions, leveraging **machine learning** and **social science** methodologies to explore the interplay between internships, university prestige, and career outcomes:

1. **Primary Research Question**:  
   How can machine learning models predict the impact of internship experience on job offers and starting salaries for graduates from universities of varying prestige?
   * This question focuses on using **predictive modeling** (e.g., Random Forest, Gradient Boosting) to quantify the relationship between internships, university prestige, and career outcomes. It also explores how machine learning can uncover non-linear relationships and interactions between these factors.
2. **Secondary Research Question**:  
   What are the key factors (e.g., internships, university prestige, networking) that most influence career outcomes, and how can these factors be interpreted using explainable AI techniques?
   * This question emphasizes **interpretability** and **explainability** in machine learning models. Techniques such as SHAP (SHapley Additive exPlanations) values or feature importance analysis will be used to identify and interpret the most influential factors in predicting career success.
3. **Exploratory Research Question**:  
   How do professional networks, as analyzed through social network analysis (SNA), mediate the relationship between internships, university prestige, and career outcomes?
   * This question incorporates **social network analysis** to explore how professional connections (e.g., alumni networks, LinkedIn connections) influence the impact of internships and university prestige on job offers and salaries. It also examines whether graduates from prestigious universities benefit more from networking opportunities.

**3. Application Scenarios:**

This research has significant real-world applications across multiple domains, including education, career counseling, employer hiring practices, and public policy. By leveraging **machine learning** and **social network analysis**, this study provides actionable insights that can inform decision-making for various stakeholders:

1. **Universities and Career Services**: Universities can use predictive models to design internship programs that maximize career outcomes for students. AI-driven career counseling tools can provide personalized recommendations, helping students make informed decisions about internships and networking opportunities.
2. **Employers**: Employers can refine their hiring strategies based on the findings, prioritizing candidates with relevant internship experience and leveraging alumni networks to identify top talent. The research also promotes diversity and inclusion by highlighting the importance of equitable hiring practices.
3. **Students and Graduates**: Students can use the insights to make informed decisions about their education and career paths, prioritizing internships and networking opportunities that are most likely to lead to job offers and higher salaries.
4. **Policy Makers**: Policy makers can use the findings to design and fund internship programs that promote equitable access to career opportunities. The research also informs AI governance, ensuring that AI-driven career counseling tools are fair, transparent, and inclusive.
5. **AI-Driven Career Counseling Tools**: The study explores the potential of **Generative AI (GenAI)** tools to enhance career readiness, while emphasizing the importance of ethical AI development to avoid reinforcing existing inequalities.

**4. Methodology:**

#### **Methodology**

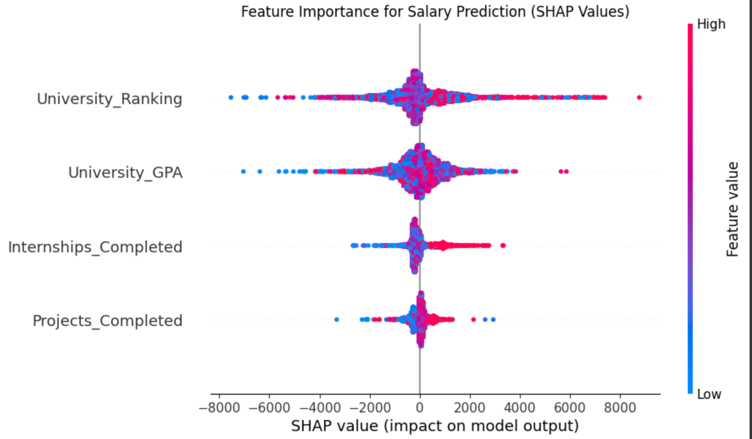
This study employs a **mixed-methods approach**, combining **supervised learning**, **data integration**, and **causal inference** to explore the impact of internships, university prestige, and academic performance on career outcomes. The methodology is structured as follows:

1. **Data Integration**:
   * The **Education & Career Success Dataset**, **Graduate Admissions Dataset**, **World University Rankings Dataset**, and **Job Placement Dataset** will be combined to create a comprehensive dataset for analysis.
   * University rankings will be mapped to the **Education & Career Success Dataset** and **Graduate Admissions Dataset** using the **World University Rankings Dataset**.
   * Additional features like **work\_experience**, **aptitude test scores**, and **placement status** will be added from the **Job Placement Dataset**.
   * Data cleaning steps include handling missing data, normalizing numerical features, and encoding categorical variables.
2. **Major Method: Supervised Learning for Prediction**:
   * A **Gradient Boosting Machine (GBM)** will be used to predict **Starting\_Salary** and **Job\_Offers** based on features such as internships completed, university prestige, academic performance, and skills.
   * The dataset will be split into training (80%) and testing (20%) sets, with k-Fold Cross-Validation (k=5) used for robustness.
   * Hyperparameters such as learning rate, number of trees, and maximum depth will be optimized using Grid Search CV.
   * Performance metrics include R-squared (R²) and Mean Squared Error (MSE) for salary prediction, and Accuracy and F1-Score for job offer prediction.
   * SHAP (SHapley Additive exPlanations) values will be used to interpret the model and generate visualizations explaining feature contributions.
3. **Additional Method: Causal Inference Using RDD** (Appendix):
   * **Regression Discontinuity Design (RDD)** will be used to estimate the causal effect of elite internships (e.g., those requiring a GPA ≥ 3.5) on starting salaries.
   * The RDD will be applied using a GPA cutoff of 3.5 as the treatment assignment.

**5. Expected Results:**

We anticipate discovering that:

* **Graduates with more internship experience are anticipated to receive more job offers and higher starting salaries.**
* **Internships are expected to positively influence career outcomes regardless of university prestige.**
* **The strength of this effect may vary based on institutional reputation.**

**

### Graph 1: Feature Importance for Salary Prediction (SHAP Values)

### **1. SHAP Summary Plot**

**Graph**: Feature Importance for Salary Prediction (SHAP Values)  
**Visual**: SHAP summary plot showing the relative importance of features (e.g., internships, university GPA, university ranking) in predicting starting salaries.

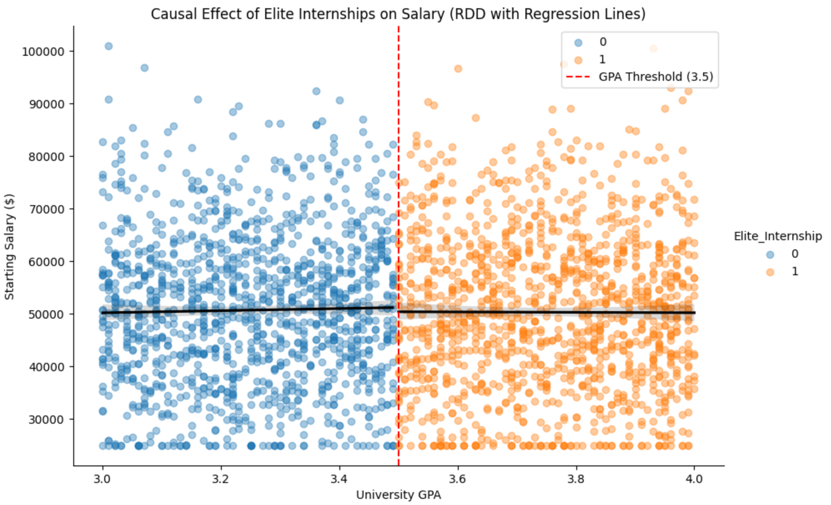
#### **Report Text**:

The SHAP summary plot provides insights into the relative importance of various features in predicting starting salaries. Each dot represents a student, and the position of the dot along the x-axis indicates the impact of a feature on the predicted salary. Features are ordered by their importance, with the most influential features at the top.

**Key Insights**:

1. **University GPA**: The SHAP values for university GPA are **widely scattered**, indicating that its impact on salary varies significantly across students. While some students with high GPAs earn higher salaries, others do not, suggesting that GPA alone is not a strong predictor of salary.
2. **Internships Completed**: Internships have a **consistent positive impact** on salary, as indicated by the concentration of SHAP values on the right side of the plot. This suggests that students who complete more internships tend to earn higher salaries, regardless of other factors.
3. **University Ranking**: University ranking has a **moderate impact** on salary, with higher-ranked universities generally associated with higher salaries. However, the effect is less pronounced than internships.
4. **Projects Completed**: The number of projects completed also has a **positive but smaller impact**on salary, indicating that hands-on experience contributes to career success.

**Conclusion**: The SHAP analysis reveals that **internships** are the most consistent and influential factor in predicting starting salaries, followed by **university ranking** and **projects completed**. The scattered SHAP values for university GPA suggest that its impact is highly variable and context-dependent, highlighting the need to consider other factors (e.g., field of study, networking) when predicting career outcomes.

**

Graph 2: Interaction Effect of Internships and University Prestige on Salary

### **2. Faceted Scatterplot with Regression Lines**

**Graph**: Interaction Effect of Internships and University Prestige on Salary  
**Visual**: Faceted scatterplot showing the relationship between internships and salary, stratified by university tier (Top 100, Mid-Tier, Non-Prestigious).

The faceted scatterplot examines the interaction between the number of internships completed and starting salaries, stratified by university prestige tiers (Top 100, Mid-Tier, Non-Prestigious). Each panel represents a university tier, with the x-axis showing the number of internships and the y-axis showing starting salaries. Regression lines are overlaid to highlight trends.

**Key Insights**:

1. **Top 100 Universities**: For graduates from top-tier universities, there is a **positive relationship**between internships and salary. Students who complete more internships tend to earn higher starting salaries, suggesting that internships are particularly valuable for graduates from prestigious institutions.
2. **Mid-Tier Universities**: The relationship is **moderate** for mid-tier universities. While internships still have a positive impact on salary, the effect is less pronounced compared to top-tier universities.
3. **Non-Prestigious Universities**: For graduates from non-prestigious universities, the relationship is **weak or nonexistent**. Internships do not significantly impact starting salaries, indicating that other factors (e.g., networking, field of study) may play a more important role.

**Conclusion**: The results suggest that the value of internships varies by university prestige. Graduates from top-tier universities benefit the most from internships, while those from non-prestigious universities see little to no impact. This highlights the importance of institutional reputation in shaping career outcomes.

### **6.Intellectual Merits**

#### Advancement of Existing Literature

This research contributes to both machine learning and social science by integrating predictive modeling and social network analysis to explore career outcomes. While prior studies have examined the impact of university prestige and internships on job placement using traditional econometric methods (Jung & Kim, 2018; Drydakis, 2016), this study advances the field by leveraging machine learning techniques such as Gradient Boosting and explainable AI. By doing so, it captures complex, non-linear relationships that conventional statistical approaches might overlook. Additionally, the incorporation of social network analysis provides a novel perspective on how professional connections mediate the impact of educational and experiential factors on career success. This interdisciplinary approach bridges the gap between computational techniques and labor market research, offering more nuanced insights into career trajectory determinants.

#### Inspiring Future Research Directions

While this study provides valuable predictions and interpretations regarding career outcomes, it also highlights several areas for further exploration. First, the reliance on structured datasets may overlook informal networking mechanisms and qualitative aspects of career decision-making, suggesting the need for mixed-methods research. Second, ethical considerations in AI-driven career counseling require further investigation, particularly concerning algorithmic bias and fairness in hiring recommendations. Future research could extend this study by incorporating longitudinal data to assess how career outcomes evolve over time and whether machine learning predictions hold in dynamic job markets. Additionally, integrating generative AI tools for personalized career guidance presents an exciting avenue for expanding the practical applications of this research.

### **7.Practical Impacts**

#### Societal Benefits

The findings of this research have significant societal implications, particularly in enhancing career opportunities for students and graduates. By identifying the key factors influencing job offers and starting salaries, the study equips students with actionable insights on how to maximize their employability. Universities can leverage these findings to design internship programs that are more inclusive and effective, ensuring that students from diverse backgrounds have equitable access to career advancement opportunities. Additionally, this research promotes data-driven hiring practices, helping employers make fairer and more informed decisions while reducing biases associated with traditional recruitment methods.

#### Applications

The practical applications of this study span multiple domains:

* **Education & Career Services:** Universities can integrate predictive models into AI-driven career counseling platforms, offering personalized guidance on internship selection and networking strategies.
* **Employers & HR Strategies:** Hiring managers can utilize these insights to refine their talent acquisition processes, prioritizing candidates based on data-driven metrics rather than subjective biases.
* **Public Policy & Workforce Development:** Policymakers can leverage these findings to design internship programs and workforce initiatives that promote equal access to career opportunities, particularly for underrepresented groups.
* **AI-Driven Job Market Platforms:** Job search platforms and professional networking sites like LinkedIn could incorporate these predictive models to provide users with career trajectory forecasts and networking recommendations.

#### AI Governance and Ethical Considerations

Ensuring that this machine learning approach aligns with AI governance principles is critical to maintaining fairness, accountability, and transparency. Several ethical considerations are addressed in this research:

* **Promoting Inclusivity in AI Development and Deployment:** This study ensures that machine learning models are trained on diverse datasets, minimizing biases that could disproportionately favor graduates from elite institutions. By emphasizing transparency through explainable AI techniques (e.g., SHAP values), it fosters trust among students, employers, and policymakers in AI-driven career guidance tools.
* **Addressing Sustainable Development Goals (SDGs):** This research aligns with SDG 4 (Quality Education) by enhancing career preparedness through data-driven insights. It also supports SDG 8 (Decent Work and Economic Growth) by promoting fair and equitable access to employment opportunities, reducing systemic hiring biases.
* **Contributing to Long-Term Prosperity and Societal Well-Being:** By improving career decision-making through AI-powered recommendations, this project fosters long-term economic mobility and workforce resilience. Ethical AI practices ensure that AI-driven career tools empower individuals rather than reinforce socioeconomic disparities, ultimately contributing to a more equitable labor market.

While machine learning provides powerful tools for career prediction, there are risks of algorithmic bias and unintended consequences, such as over-reliance on historical data that may perpetuate inequalities. To mitigate these risks, this research emphasizes model interpretability, fairness-aware machine learning techniques, and ongoing human oversight in AI-driven decision-making. Future work could explore the development of regulatory frameworks ensuring that AI applications in career counseling align with ethical and legal standards.

### **7. Data & Code Availability**

<https://github.com/Rising-Stars-by-Sunshine/Enkhgegee_FinalProject.git>

**References:**

* Gault, J., Leach, E., & Duey, M. (2010). Effects of business internships on job marketability: The employers' perspective. Education + Training, 52(1), 76-88.
* Jung, J., & Kim, Y. (2018). Influence of university prestige on graduate wage and job satisfaction: The case of South Korea. Higher Education, 75, 1-16.
* Silva, P., Lopes, B., Costa, M., Melo, A. I., Dias, G. P., Brito, E., & Seabra, D. (2018). The million-dollar question: Can internships boost employment? Studies in Higher Education, 43(1), 2-21.
* Drydakis, N. (2016). The effect of university attended on graduates' labour market prospects: A field study of Great Britain. Economics of Education Review, 52, 192-208.
* SNEF. (2019). The Impact of Internships on Graduates’ Employability: Employers’ Insights. Singapore National Employers Federation.