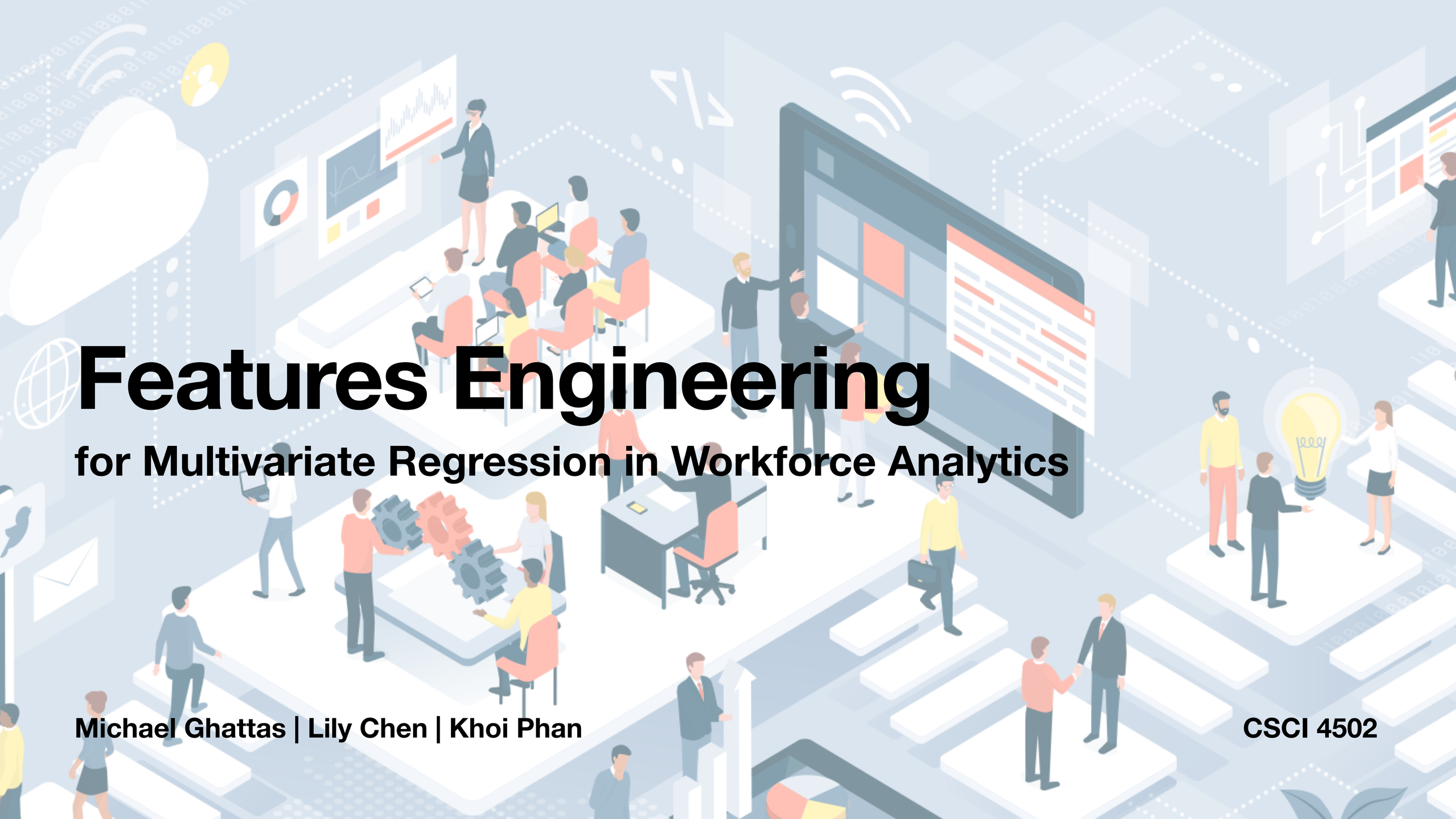


Features Engineering

for Multivariate Regression in Workforce Analytics

Michael Ghattas | Lily Chen | Khoi Phan

CSCI 4502

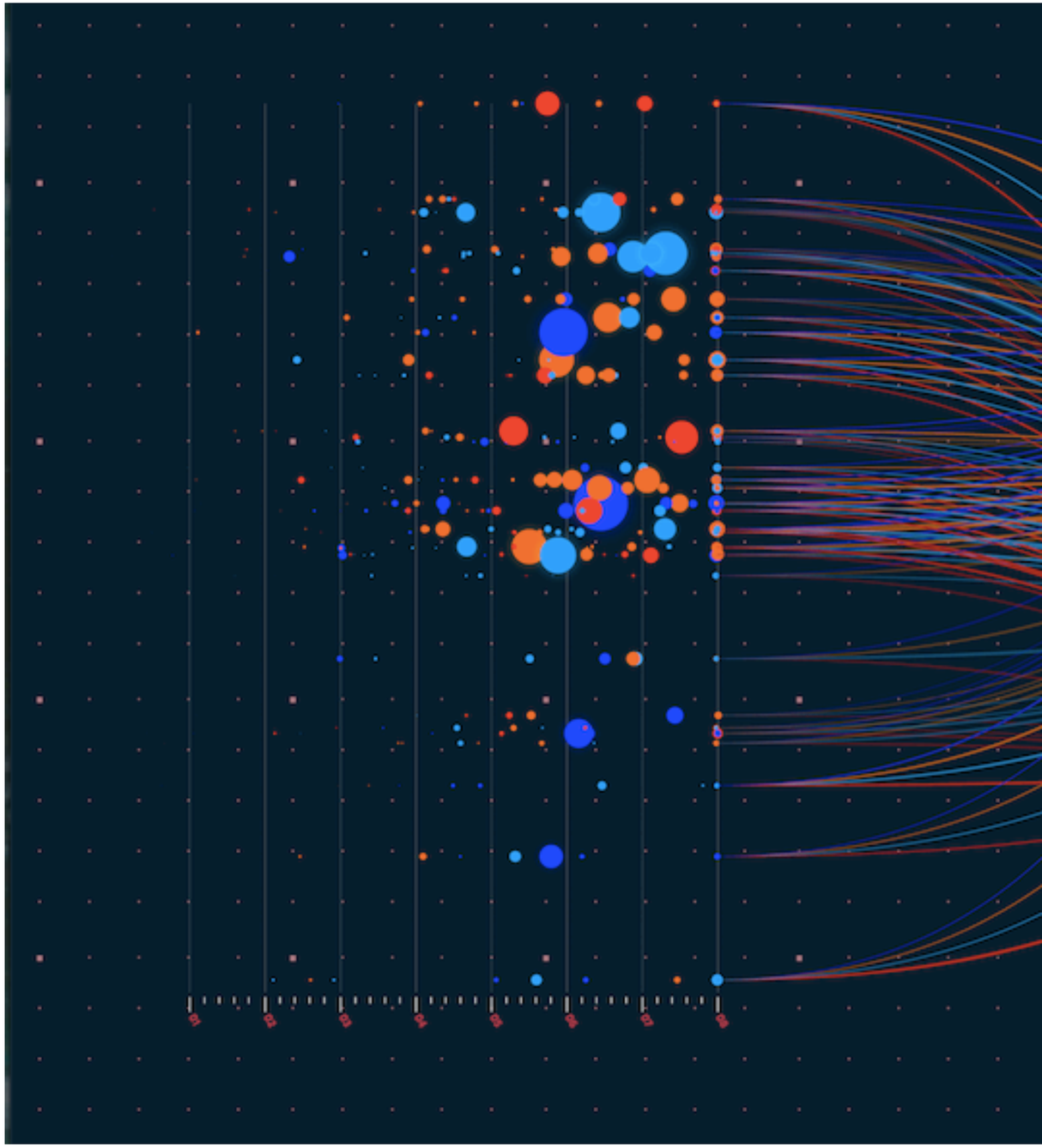
Background

- Data science growing dominance in the business world
- ROI growth mindset
- Human resources/capital optimization and management
- Social science of organizational behavior
- Inherent bias within the realm of social science
- Building models that can serve an organization’s goals

Project Description

The project aims to develop a methodology for a supervised ML algorithm utilizing a well-defined series of techniques.

Additionally, the goal is to generate the needed predictors from existing data that can serve a Multivariate Linear Regression Model aimed at measuring candidate/ employee potential.



Data Sets

https://www.aihr.com/blog/hr-data-sets-people-analytics/

- Multiple data-sets with varying information on people analytics
- Job Classifications
- Employee Engagement
- Absenteeism at Work
- Employee Attrition and Performance
- Turnover Data

Tools

R | Python

Programing Languages:

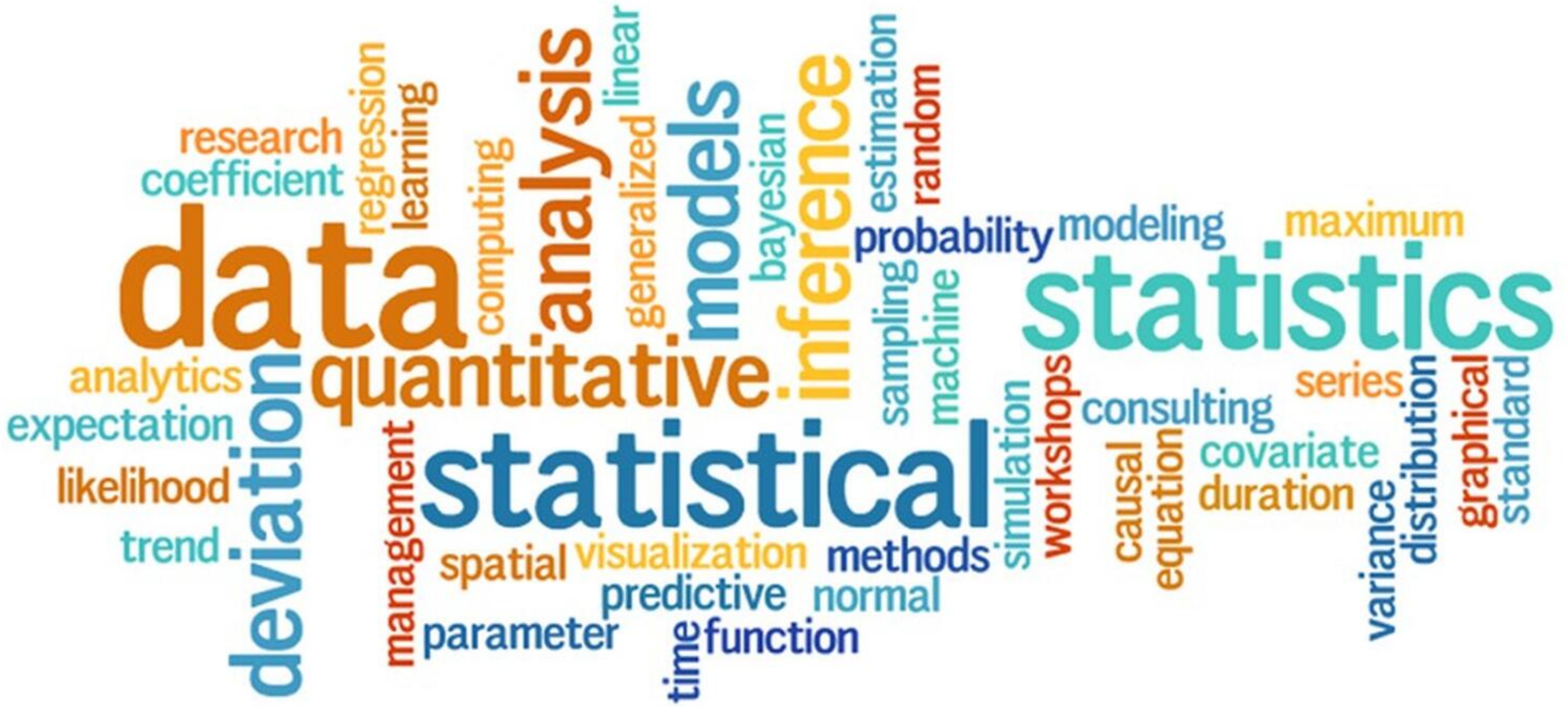
- R
- Python

Software Applications:

- JupyterLab
- VSCode

Possible Methods:

- Dynamic Discretization
- Clustering
- Deviation
- Multidimensional Analysis
- Imbalance Ratio
- Hypothesis Testing
- Akaike information criterion (AIC)
- Bayesian information criterion (BIC)
- K - Nearest Neighbor (KNN)
- Monte-Carlo Markov Chains (MCMC)



Thank you!

Please let us know if you have any questions.

Michael Ghattas | Lily Chen | Khoi Phan

CSCI 4502