IOSR Journal Of Humanities and Social Science (IOSR-JHSS)
Volume 30, Issue 9, Series 5 (September, 2025) 18-24
e-ISSN: 2279-0837, p-ISSN: 2279-0845.
www.iosrjournals.org

Data-Driven Career Path Modelling For Information Designers In Pearl River Delta: A BOSS Workforce Analytics Approach

Xing Yang
School Of Arts And Design, Huizhou University, Huizhou, China

Lei Qiao*
School Of Creativity And Design, Guangzhou Huashang College, Guangzhou, China

Xinzhe Guan
Faculty Of Humanities And Arts, Macau University Of Science And Technology, Macau, China

Kun Xing
Academy Of Art And Design, Anhui University Of Technology, Maanshan, China

Abstract
The career of information design is becoming a strategic one to facilitate data-driven visual communication, user experience design, and data storytelling in all industries of an economy anchored on data. In the current study, to model the career paths of information designers working in the Pearl River Delta (PRD) region, the combination of the Markov Chain Modelling concept with labour market data on the BOSS Zhipin platform will be used. The study generates clean data sets by scraping and parsing job advertisements with Natural Language Processing (NLP) to restructure job advertisements unstructured data into a structured form of the job titles, skill requirements, industry samples, and salary range. Probable career transitions are then simulated using the Markov Chain Model, the probability of state-to-state movements is estimated and the most stable and volatile career stages determined. The findings denoted the existence of high probabilities of occupational mobility between entry quality jobs, including Junior UX Designer or Visual Analyst and a mid-level occupation being UI/UX Specialist or Information Architect. This paper provides a repeatable procedure of using Markov Chains in modelling of creative-technical careers and reveals the promise of workforce analytics to maximise utilisation of human capital within the innovation ecosystems in regions.

Key Words*: Information Design, Career Path Modelling, Markov Chain, Workforce Analytics, Pearl River Delta*

Date of Submission: 16-09-2025 Date of Acceptance: 22-09-2025

I. Introduction

Amidst the current data-obsessed world, information designers have become a key position between data analysis, visual narrative, and user-friendly communication[1]. These experts convert complex data sets and present them in easy to read and to use visuals and interfaces that inform decision making in industries[2]. The need of professionals skilled in information design has soared especially in innovation-intensive areas such as the PRD), whose economic developments rely on digital infrastructure. Nevertheless, the professional area is still uneven in relation to their career path, and empirical data on realizing their duties, their extension in various industries, and their future development greatly lack[3].

To fill this gap, this paper suggests a data-based career path modelling solution that integrates labour market insights of the BOSS Zhipin recruitment firm and the probabilistic Markov Chain Modelling. Based on methods augmenting and structuring data provided on job ads using NLP in job title, skills, industry and salary data, the study builds transition matrices to model the careers under analysis and define exceptionally stable versus wild jobs[4]. This framework also covers the actual mobility trends of the information design industry in the PRD but also determines the workforce development, educational alignment, and talent strategy. The methodology herein tries to maximize the strategic planning of human capital in the creative-technical disciplines, particularly within vibrant innovation systems[5].

II. Literature Review

In order to frame the proposed approach, the previous literature is analyzed on modelling career paths and workforce analytics and the application of statistical, machine, learning, and NLP modelling techniques. An overview of important contributions, methodologies tested, and their advantages as well as limitations are presented in Table 1.

Table 1: Key literature on workforce and career modelling

Author(s)	Techniques	Advantages	Limitations
C. Qu and E. Kim[6]	AI-enabled innovation framework	Supports MSME innovation	No career path modelling
L. Zhang[7]	Data-driven HR planning	Strategic workforce alignment	Lacks individual transition focus
W. Yu et al.,[8]	Digital resource bundling	Enhances supply chain adaptability	Not relevant to design roles
R. Quilbé et al.[9]	Policy & field implementation analysis	Provides real-world context	Outside workforce domain
Y. Sun et al.,[10]	Job behaviour analytics	Tracks labour shifts in real-time	No micro-level career modeling

Although the literature examined has given us ideas on workforce planning and the innovation ecosystems, it has not put much focus on how the individuals make career transitions on a micro level particularly in creative-technical disciplines such as information design. Majority emphasise on narrating analysis at a strategic or policy level process naturally involving graphic models of individual role mobility or hybrid job dynamics. In order to fill these gaps, this paper will apply Markov Chain model to the real-time job data of BOSS Zhipin, extracted with NLP, and pulled structured career features. The specified strategy simulates transitions, the determination of solid and turbulent roles, and the orientation of the labour trend into education and workforce planning of the innovation-based economy in the Pearl River Delta.

III. Career Transition System For Information Designers

The proposed workflow of this study aimed at modelling career paths of information designers in PRD-region will adhere to the logical five-step strategy to convert the unorganized data on labour markets into viable workforce information. The given process starts with the stage of data acquisition when job advertisements will be scraped on the BOSS Zhipin platform to get detailed data about job titles, necessary skills, industries, salaries, etc. Secondly, under NLP and structuring step, the raw text-based information gets cleansed and even processed through tokenization, NER, and Part-of-speech tagging, so as the key attributes are identified and structured into a format. The third stage consists of job classification and career state definition: the normalisation of similar position structures, grouping of skills according to theme, and the binning of salary levels - the creation results in clearly defined states of careers, the nodes of which will form the basis of a Markov Chain. In the figure 1, the architecture needed was as proposed.

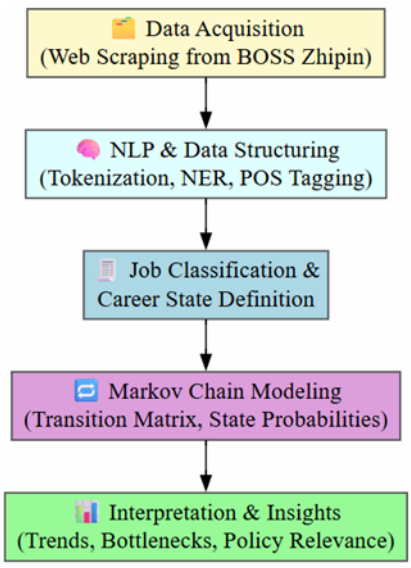

Figure 1: Proposed system architecture

The fourth step undertakes Markov Chain modelling where a transition probability matrix is created to model career transition across occupational states basing on the historical rate of role changes, skills overlap and using the probability of a worker to be a certain state under that occupational condition. The usage of this matrix allows the identification of steady and volatile states in career trajectories. Lastly, during the interpretation stage, the outputs of the model are viewed in order to identify the patterns of occupational mobility, the

pathways of skills development and bottlenecks in the process of role performance. The findings will be placed within the context of providing strategic directions to educators, employers and policy-makers with a vision of bringing the right talent onto the path of advancing the digital and innovation-led economy in the PRD.

Data Acquisition via Web Scraping

In the initial phase of the suggested architecture, the goal is to pull information on job advertisements available through BOSS Zhipin, the largest Chinese recruitment site that lists many job opportunities in design and digital fields. Job-specific web scraping needs a programmed web scraper, say, using Scrapy or Selenium, honed to harvest the correct information, respectively, after JavaScript and/or xHTML data. The most crucial items of data are job titles, skills required, remunerations, nature of corporations, industrial sectors, experience levels, and regions. This is to give the information design roles wide coverage on sectors in the PRD region[11].

In order to keep the data quality and keep it relevant, the filters are used to remove copied data, outdated posts, and irrelevant types of jobs not reflecting the creative-technical roles. The scraper is structured to deal with dynamic contents of the page and with pagination of results to optimise data-retrieval. The result of the step is a clear, unstructured data that mirrors current labour market situation. Such information is the fundamental data of the following NLP procedure, which can lead to the systematic exploration of career patterns in the field of information design.

Data Structuring and Natural Language Processing

During this phase, raw data on job adverts is cleaned to a structured and analyse form by data enrichment using NLP techniques. Initial steps in the text processing procedure merely entail tokenization (dividing a text into individual words or groups of words), cutting of stop words (words that are normal and do not add any meaning like the and the) and normalization (formatting all the text to one common form). These back-of-the-envelope manipulations clean the data and make it ready to be most closely examined at the semantic level.

Further NLP operations like NER and POS tagging are then used in extracting sensible entities in the text- e.g. job titles, skills, salary levels, experience desired and the company names. The extracted elements are grouped into labelling and mapped into standardized labelling's. As an example, we may list such a role like the Visual Interface Designer and the UI Designer brought to a common role type. This ordered data is then put in a database framework, giving an easy way of categorizing it and tagging it uniformly in terms of job roles, skill sets, etc. This is paramount to the accuracy of the modelling of career path later on[12].

Job Classification and State of Career Definition

Once the raw data has been organized into a structure it is the process of identifying career states by organizing the data intelligibly into professional states that should be defined into career state configuration that can be utilized in the process of modelling professional transition. The results obtained by extracting job titles using NLP are normalized by grouping the similar roles into single groupings. To give an example, such job titles as "Junior UX Designer," "UX Intern," and "User Experience Assistant" are placed in one consolidated entry-level category. Mid-graduate and top positions, e.g., the ones labelled as "UI/UX Specialist" or "Information Architect" are also lumped together. This normative clarifies that the various companies and job advertisements have the same level of labelling the various career levels.

Then, skills are clustered with regard to topics (e.g. design tools, frontend coding, research methods), and the salary ranges are discretized into bins (e.g. low, medium, high). The set of job title, skill category and salary tier characteristics uniquely mean a career state; each career state is a node in the Markov Chain. The whole job environment is therefore modelled as a collection of discrete and connected states where each s corresponds to a particular career phase. Determination of these states is used in determining the transition probabilities in the subsequent procedure and so is the ability to simulate how the professionals tend to shift across roles as time progresses in the world of information design.

Markov Chain Modelling for Career Transitions

At this stage the Markov Chain Model is used to simulate and investigate the career progression patterns of information designers. A transition matrix is created using the career states identified in Step 3 to reflected the probability of going one occupational state to another in a minimum of one phase of time. Every element of the matrix indicates the rate at which a job seeker moves between two particular jobs-in terms of the resemblance of skills, progression in pay, step-wise description of the job outlined in the data. As an example, in case the transitions between the states of Junior UX Designer and UI/UX Specialist are frequent, the probability of their transition between the states would also be high. Mathematically the Markov Chain is shown using a transition probability matrix P with each element representing the probability of making a transition between state and state whereby equation (1),

(1)

The model is Markovian and, that is, the consecutive state is given by the present state and not by the preceding series of steps. After the matrix has been obtained; the analysis of the same is carried to determine a steady-state probabilities-which are states that workers will remain in over the time-and volatile states that have high turnover. Such visualizations display such insights to represent common career tracks found within the information design sector of the PRD one of which being an upward advancement between the job titles of Visual Analyst and Information Architect with the idea of assisting stakeholders to comprehend how the workforce dynamics occur within the information design sector of the PRD.

Interpretation and Workforce Insights

In the last step, actionable findings concerning the pattern of career development within the information design career are made with the help of the outputs of the Markov Chain model. The transition probabilities between the various occupational states display the important mobility patterns, including those of frequently upward moves (e.g, from a Junior UX Designer occupation to the Information Architect one) and places of career bottlenecks where people will tend to stay without further advances. These trends are useful to determine those occupations that act as stepping stones and other positions that might need new skills or company resources to further their chances. Moreover, career road maps are created in order to see how certain skills (e.g. UI prototyping, user research etc.) will affect the career development[13].

The insights are put into perspective within the framework of the overall innovation ecology of the PRD, in which the information designers play an important role of providing digital transformation user-centered solutions, data storytelling. The study, through an analysis of labour mobility in this industry, makes some good suggestions on how education institutions can make adjustments based on labour market needs and employers to make a more efficient design of career ladders and training plans. Moreover, policymakers will be able to utilise such findings to promote the workforce development strategies that meet the specifics of the digital economy that develops in the region and provides the best possible use of human resources in creative-technical field.

IV. Findings

This part contains and explains the most important findings of the proposed career path modelling system. Using the Markov Chain analysis on well-structured job advertising data, the research finds out that certain transition tendencies are shared, that there is stability and volatility in career states, and the upward mobility trends in the area of information designers within the PRD region. Those findings can be viewed against the concepts of skill building, employment market, and workforce strategy and can provide great insights to practitioners, educators, and policymakers.

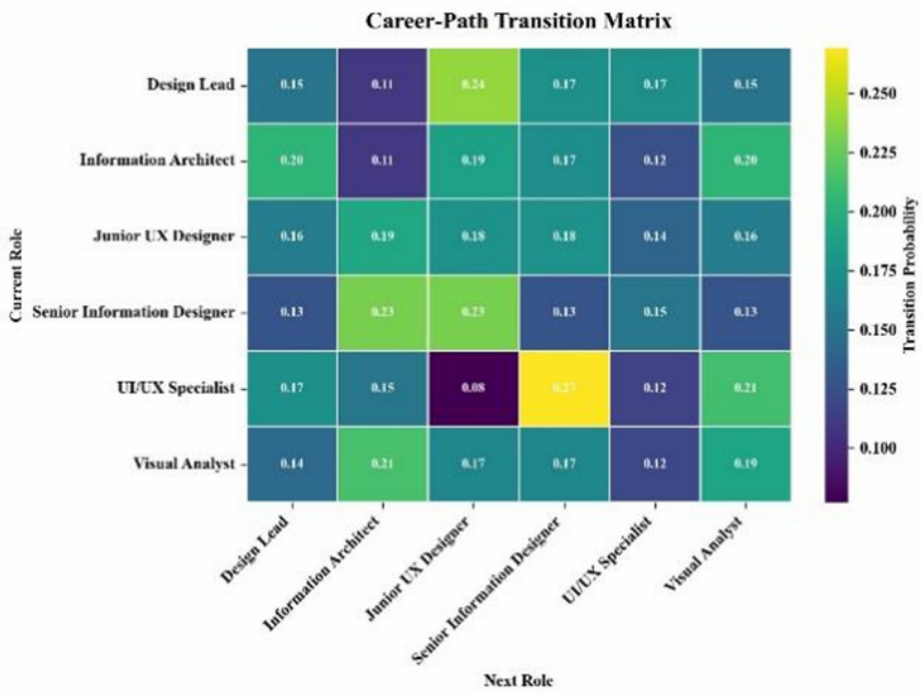

Figure 2: Career Path transition matrix

Figure 2 is the career-path transition matrix and shows the probabilities of changes in roles within the information designers. There is no downward trend and a high transition probability (0.27) has been found between UI/UX Specialist and Senior Information Designer. The transition probabilities between Senior Information Designers and Information Architect or Junior UX Designer are equal (0.23), which indicates the ability to switch between the specializations. Design Leads become most commonly Junior UX Designer (0.24), and Visual Analysts turn into Information Architect (0.21). In the same way, Information Architects also shift to Design Lead and Visual Analyst (0.20 each). These values point out the dynamism in career mobility, identified with evident trends in vertical channels as well as horizontal movement of various roles within the information design profession.

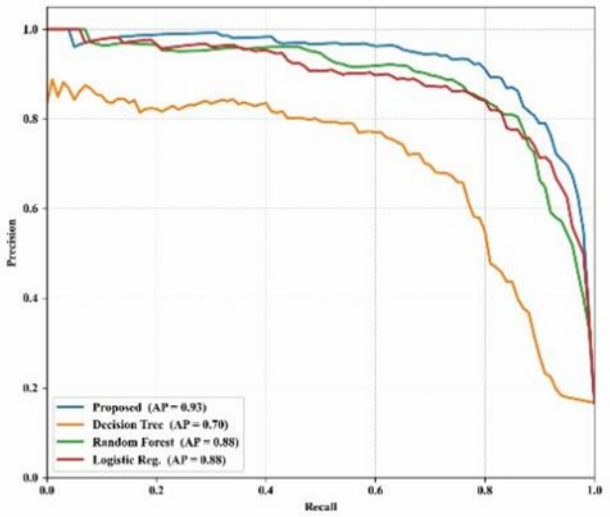

Figure 3: Precision and recall analysis

Figure 3 presents Precision-Recall (PR) curves of the evaluation of the performance of the proposed model along with Decision Tree, Random Forest, and Logistic Regression classifiers. The presented model surpasses all baselines with the best average precision (AP) of 0.93 which shows a better trade-off of precision and recall. Random Forest and Logistic Regression models come next with AP of 0.88 each which is a comparatively good but not the brightest performance. The Decision Tree model is the worst one with a clear decrease in the precision in the recall range and AP of 0.70. The outcomes reveal that the suggested system better decreases the ideas of relevant changes in a career path modelling with fewer false positives.

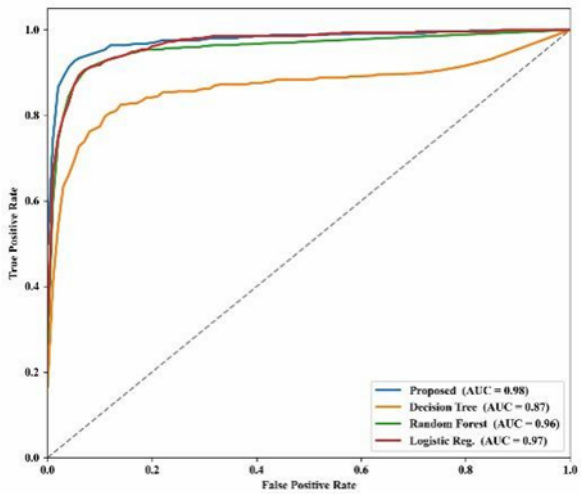

Figure 4: AUC validation

Figure 4 shows the ROC curves of models in terms of the classification results of the proposed model in comparison to that of Decision Tree, Random Forest, and Logistic Regression classifiers. The result of the proposed model provides the Area Under the Curve (AUC) of 0.98, which is excellent in differentiating the true and false positives. Logistic Regression takes the next place with an AUC of 0.97 and Random Forest with 0.96,

which are also the ones that show significantly high predictions. Decision Tree is already slacker and presents an AUC of 0.87, indicating the relatively worse performance. The steeper and left-side oriented curve of proposed model proves the effectiveness of the factor to identify the correct transitions between career paths with the least false alarms and further the proof of the excellence in entire system.

V. Discussion

The proposed career path modelling system demonstrates strong analytical capabilities in identifying mobility trends within the information design profession in the PRD region. Markov Chain analysis reveals significant upward transitions, notably a 0.27 probability from UI/UX Specialist to Senior Information Designer, and balanced horizontal mobility with transition probabilities of 0.23 between Senior Information Designer and both Information Architect and Junior UX Designer. The system's classification performance is validated through precision-recall analysis, where it achieves the highest average precision (AP) of 0.93, outperforming Random Forest and Logistic Regression (both at 0.88), and significantly ahead of Decision Tree (0.70). Furthermore, the ROC-AUC evaluation confirms the robustness of the model, with an AUC of 0.98, compared to 0.97 (Logistic Regression), 0.96 (Random Forest), and 0.87 (Decision Tree). These results underscore the model's ability to detect accurate career transitions with minimal false positives. The identified career dynamics provide practical insights into role fluidity and promotion patterns. Overall, the system supports strategic decision-making for workforce development, skill-building, and policy interventions.

VI. Conclusion

The paper has outlined an empirical approach of modelling career shifts in the information design industry in the PRD region by combining the real-time data at labour market in BOSS Zhipin with Natural Language Processing and transition modelling based on Markov Chain Modelling. The data shows evident movements with entry-level workers shifting to mid-level job positions, notably Junior UX Designer and Visual Analyst to UI/UX Specialist and Information Architect. Although it shows the potential of workforce analytics in dynamic flow of career capturing, the model is subject to limitations of its dependency on a single job platform as well as the exclusion of individual career history records and current employer need requirements in real-time. The next study has the potential to add more substantial value to the model by including the multi-data source, career path tracking over possibly time, and a hybrid learning approach to designing a career planning tool closer to personal predictive planning. Correction of such shortcomings, the study can be used to provide a scalable and repeatable process of grasping and maximizing talent development in innovation-based economies.

Acknowledgement

The study is supported by Guangdong Province 2023 Education Science Planning Project (Higher Education Project), "Information Design Talent Demand and Career Pathways in the Pearl River Delta" (No.: 2023GXJK508); 2023 Guangdong Province Philosophy and Social Sciences Planning Co-construction Project, "Research on the Redesign of Digital Technology to Revitalize Endangered Pixiu Dance Culture" (No.: GD23XYS050); Guangzhou Huashang College 2023 Undergraduate University Teaching Quality and Reform Project (First-class Major Project), "Visual Communication Design" (No.: HS2023ZLGC01).

References

[1]. P. Niu, J. Zhu, And Y. Sun, "Dynamic Modeling And Chaos Control Of Informatization Development In Manufacturing Enterprises," Entropy, Vol. 23, No. 6, P. 681, 2021.
[2]. P. Kun, M. Esa, And P. Munianday, "Optimizing Construction Cost Control Through Bim5d Digitalization: A Case Study On The Application Of Glodon Software In China," J. Adv. Res. Appl. Sci. Eng. Technol., Vol. 32, No. 2, Pp. 487–507, 2023.
[3]. S. Park, A. Tafti, And G. Shmueli, "Transporting Causal Effects Across Populations Using Structural Causal Modeling: An Illustration To Work-From-Home Productivity," Inf. Syst. Res., Vol. 35, No. 2, Pp. 686–705, 2024.
[4]. S. Liu, R. Thurasamy, And S. R. H. Hati, "Determinants And Outcomes Of Green Technology Innovation Adoption Among Third-Party Logistics Firms In China: A Sem-Ann Analysis," Systems, Vol. 12, No. 9, P. 331, 2024.
[5]. A. Kuang, G. Yeung, M. Wang, And Y. Tong, "Digital Platform, Spatial-Digital Fix, And The Reconfiguration Of Apparel Production Networks," Econ. Geogr., Vol. 100, No. 5–6, Pp. 381–409, 2024.
[6]. C. Qu And E. Kim, "Artificial-Intelligence-Enabled Innovation Ecosystems: A Novel Triple-Layer Framework For Micro, Small, And Medium-Sized Enterprises In The Chinese Apparel-Manufacturing Industry," Sustainability, Vol. 17, No. 11, P. 5019, 2025.
[7]. L. Zhang, "Data-Driven Decision Support Framework For Human Resource Strategic Planning," Int. J. High Speed Electron. Syst., Vol. 33, No. 1, P. 2540197, 2024.
[8]. W. Yu, M. Jacobs, R. Chavez, And Y. Song, "Leveraging Supply Chain Visibility For Implementing Just-In-Case Practices: The Roles Of Knowledge And Digital Resources Bundling," Supply Chain Manag.: Int. J., Vol. 29, No. 5, Pp. 871–884, 2024.
[9]. R. Quilbé Et Al., "Cold Water Thermal Refuge Enhancement And Creation For Salmonids: Successes, Failures, And Lessons Learned," River Res. Appl., 2025, Early Access.
[10]. Y. Sun, H. Zhu, L. Wang, L. Zhang, And H. Xiong, "Large-Scale Online Job Search Behaviors Reveal Labor Market Shifts Amid Covid-19," Nature Cities, Vol. 1, No. 2, Pp. 150–163, 2024.

[11]. F. A. Bhat And S. Parvez, "Evaluating Industry 4.0's Sustainable Performance And Its Impact On The Manufacturing Strategies," Proc. Inst. Mech. Eng. Part J: J. Eng. Tribol., Vol. 238, No. 10, Pp. 1262–1272, 2024.
[12]. D. Wang, J. Wang, And Z. Xiang, "Imbrications Of It And Hospitality Organizations," Ann. Tour. Res. Empir. Insights, Vol. 2, No. 2, P. 100021, 2021.
[13]. J. S. Retzky Et Al., "Efficacy Of Transection-Free Patellar Tendon Imbrication Technique For The Correction Of Patella Alta," Orthop. J. Sports Med., Vol. 13, No. 4, P. 23259671251325752, 2025.