# Improving Career Path Prediction with Temporal Decay Weighting on an MLP-Based Sentence Transformer Pipeline

Nezar Abdilah Prakasa
*Master's Program in Artificial Intelligence, Universitas Gadjah Mada, Indonesia*

**Abstract**

Career path prediction supports internal workforce mobility by anticipating an employee's most likely next occupation from their work history. Senger et al. (2025) introduced a strong baseline that encodes an employee's entire career history with a fine-tuned sentence transformer, projects it through a multi-layer perceptron (MLP), and retrieves the closest matching occupations from the ESCO taxonomy. This pipeline, however, aggregates all past job experiences with equal weight, irrespective of how long ago they occurred. We propose temporal decay weighting, a lightweight, inference-only modification that encodes each job experience individually and combines the resulting embeddings through a weighted average in which more recent experiences receive exponentially larger weights, controlled by a decay parameter λ. Evaluated on the KARRIEREWEGE dataset under the decorte configuration, an ablation over $\lambda \in [0.1, 1.0]$ shows that $\lambda = 0.7$ raises Recall@10 from 42.56% to 44.34% (+4.18%), while $\lambda = 0.9$ yields the best MRR (25.81%, +1.49%) and Recall@5 (36.07%, +1.55%). The method requires no retraining and can be applied to any embedding-based career recommendation pipeline.

***Keywords:*** *career path prediction, temporal decay weighting, sentence transformers, internal mobility, ESCO*

## 1 Introduction

In modern organizations, internal workforce mobility , moving employees into new roles within the organization rather than relying on external hiring , has become a central pillar of talent management (Bidwell, 2011; Ray, 2024). Effective internal mobility depends on the organization's ability to anticipate which roles an employee is likely to move into next, based on their accumulated work experience (Campion et al., 1994; Call et al., 2025).

Traditionally, such decisions rely on manual review by HR professionals and line managers, a process that is time-consuming, inconsistent, and difficult to scale across large organizations (Maley et al., 2024; Zhu et al., 2023). Recent progress in natural language processing (NLP) has opened the possibility of automating career path prediction directly from textual work-history data.

Decorte et al. (2023) were among the first to formulate career path prediction as a representation-learning problem: a candidate's resume is encoded with a sentence transformer fine-tuned via contrastive learning, and the resulting embedding is matched against embeddings of occupation titles from the European Skills, Competences, Qualifications and Occupations (ESCO) taxonomy. Senger et al. (2025) extended this line of work substantially, releasing the large-scale KARRIEREWEGE dataset and proposing an MLP-based pipeline that projects a resume embedding into the embedding space of the candidate's likely next occupation, evaluated under several dataset configurations.

Despite their strong results, both pipelines share a common simplification: a candidate's entire career history , regardless of whether an experience occurred ten years ago or one year ago , is aggregated into a single embedding with uniform contribution. Senger et al. (2025) explicitly identify the lack of temporal modeling as an open limitation of their work.

In this paper we address this gap with *temporal decay weighting* (TDW), a simple modification applied at inference time, with no retraining of the underlying sentence transformer or MLP. Each job experience in a candidate's history is encoded separately, and the experience-level embeddings are combined through a weighted average governed by an exponential decay function of recency. We conduct a systematic ablation over the decay parameter λ on the KARRIEREWEGE dataset (decorte configuration), and provide an analysis of the resulting precision–coverage trade-off and its implications for HR applications. The remainder of this paper is organized as follows: Section 2 reviews related work, Section 3 describes our method, Section 4 details the experimental setup, Section 5 reports results, Section 6 discusses their implications, and Section 7 concludes.

## 2 Related Work

### *2.1 Career Path Prediction*

Decorte et al. (2023) propose a resume representation learning approach for career path prediction: resumes are encoded with a sentence transformer fine-tuned using a contrastive objective on pairs of consecutive job titles, and predictions are produced by matching the resulting embedding against an inventory of ESCO occupation embeddings. Senger et al. (2025) build on this formulation with the KARRIEREWEGE dataset, a large-scale resource of career sequences, and propose a feed-forward MLP that maps a resume embedding into the embedding space of the subsequent occupation. Under their decorte evaluation configuration , designed to mirror the original Decorte et al. (2023) split for comparability , the MLP pipeline achieves Recall@10 = 42.56%, which we adopt as our baseline.

### *2.2 Sentence Representations*

Sentence-BERT (Reimers and Gurevych, 2019) introduced a siamese network architecture for producing sentence embeddings whose cosine similarity reflects semantic similarity, building on the BERT architecture (Devlin et al., 2019). Both Decorte et al. (2023) and Senger et al. (2025) fine-tune sentence-transformer models of this family for the career-path domain. The MLP component of Senger et al.'s pipeline is itself a simple feed-forward network trained with backpropagation (Rumelhart et al., 1986), projecting between two fixed-size embedding spaces.

### *2.3 Temporal Modeling in Sequential Recommendation*

Outside the career-prediction literature, temporal information has long been recognized as informative in sequential recommendation. Chen et al. (2022) propose time-lag–aware sequential recommendation, explicitly modeling the interval between consecutive user interactions. Hassan et al. (2022) propose an exponential decay function to weight historical user interactions by recency in e-commerce recommendation, a functional form closely related to the one we adopt in Section 3.2. Oh and Cho (2024) study recency bias in sequential recommender systems, showing that models often over- or under-weight recent interactions depending on dataset characteristics. To our knowledge, none of these temporal weighting schemes have previously been applied to sentence-transformer/MLP pipelines for career path prediction; our work brings this idea to that setting.

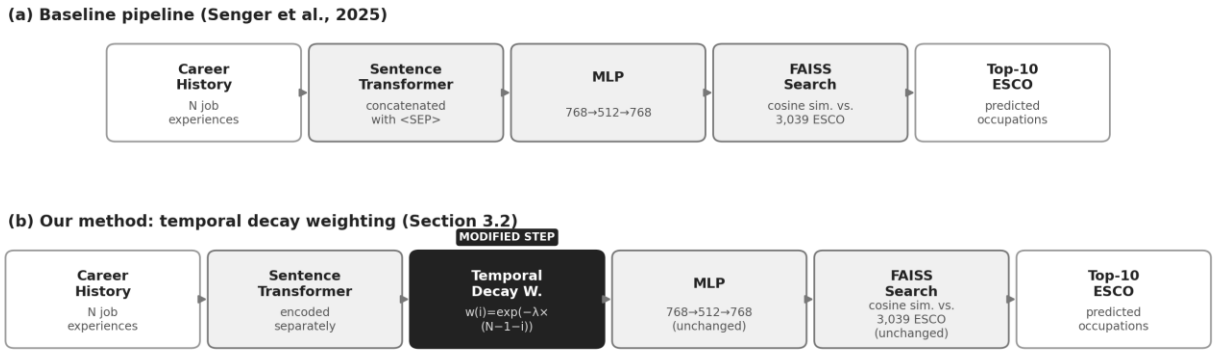


*Figure 1: Comparison of (a) the baseline pipeline (Senger et al., 2025) and (b) our temporal decay weighting modification. Only the highlighted step differs; the encoder, MLP and FAISS search are unchanged.*

## 3 Method

### *3.1 Baseline Pipeline*

We adopt the MLP pipeline of Senger et al. (2025) as our baseline without modification. Given a career history consisting of N job experiences ordered chronologically, the baseline concatenates the role titles and descriptions of all experiences into a single text, separated by a special <SEP> token. This text is encoded by a sentence-transformer model (a fine-tuned all-mpnet-base-v2 checkpoint released by Senger et al.) into a single 768-dimensional vector h. The vector h is passed through a two-layer MLP (768 → 512 → 768, with a ReLU activation), trained with a cosine embedding loss against the embedding of the ground-truth next occupation. At inference time, the MLP output is compared via cosine similarity to the embeddings of all 3,039 ESCO v1.1.2 occupation titles using a FAISS index, and the top-10 most similar occupations are returned as predictions.

### *3.2 Temporal Decay Weighting*

Our modification changes only how the input vector h to the MLP is constructed; the sentence-transformer

encoder, the MLP, and the FAISS retrieval step are all used as released, without retraining.

Rather than concatenating all N experiences into a single text, we encode each experience $i \in \{0, ..., N-1\}$ (indexed chronologically, with $i = N-1$ the most recent) separately with the same sentence-transformer encoder, yielding embeddings $e_0, ..., e_{N-1} \in R^{768}$. We then define an exponential decay weight for experience $i$:

$$w(i) = \exp(-\lambda \cdot (N - 1 - i)) \quad (1)$$

where $\lambda \geq 0$ is a decay parameter. Weights are normalized to sum to one:

$$\hat{w}(i) = w(i) / \Sigma_{j=0}^{N-1} w(j) \quad (2)$$

and the final representation passed to the MLP is the weighted average

$$h = \Sigma_{i=0}^{N-1} \hat{w}(i) \cdot e_i \quad (3)$$

When $\lambda = 0$, all weights are equal ($\hat{w}(i) = 1/N$ for all $i$), recovering a simple unweighted average. As $\lambda$ increases, the weight assigned to recent experiences grows exponentially relative to older ones.

***Worked example.*** Consider a three-experience career history (N = 3): "Chef de Cuisine" (10 years ago, $i = 0$), "Chef / General Manager" (5 years ago, $i = 1$), and "Lead Cook" (1 year ago, $i = 2$), with $\lambda = 0.7$. The unnormalized weights are w(0) = exp(−0.7 × 2) = 0.247, w(1) = exp(−0.7 × 1) = 0.497, and w(2) = exp(0) = 1.000, summing to 1.744. After normalization, ŵ(0) = 0.142, ŵ(1) = 0.285, and ŵ(2) = 0.574 , the most recent experience contributes more than four times as much as the oldest one to the final representation h.

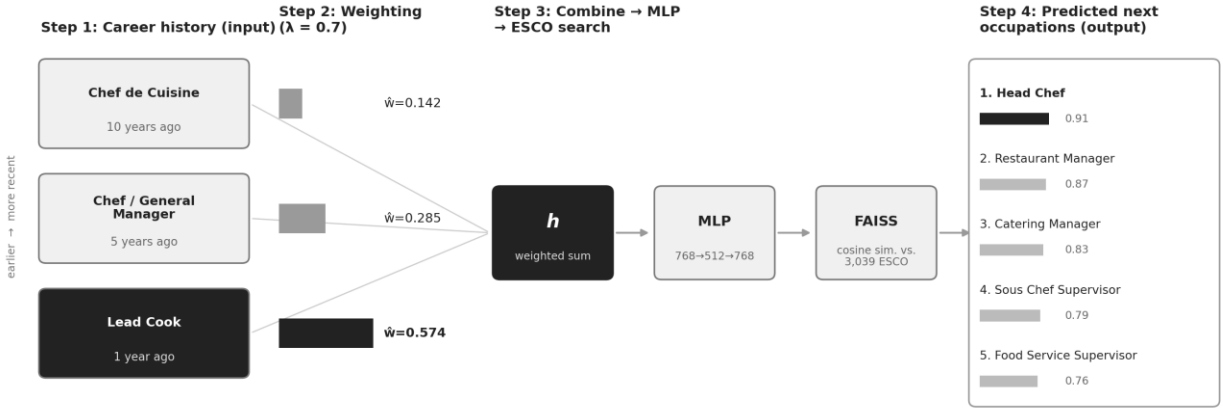


*Figure 2: End-to-end illustration for the worked example in Section 3.2 (lambda = 0.7). More recent experiences receive larger normalized weights. The representation h is passed through the unchanged MLP and FAISS search. Predicted occupations and scores are illustrative.*

## 4 Experimental Setup

### *4.1 Dataset*

We evaluate on the KARRIEREWEGE dataset (Senger et al., 2025) under its decorte configuration, which mirrors the train/validation/test split used by Decorte et al. (2023): 1,720 training, 217 validation, and 227 test career sequences. Following Decorte et al. (2023), every sub-sequence of length ≥ 2 within a career sequence is treated as an independent prediction instance, with augmentation applied after the train/validation/test split to avoid data leakage. After augmentation, the test set contains 1,802 instances. The prediction target space is ESCO v1.1.2, comprising 3,039 occupation titles.

### *4.2 Implementation Details*

We use the publicly released checkpoints from Senger et al. (2025) without modification: the fine-tuned all-mpnet-base-v2 sentence-transformer encoder and the MLP weights trained for the decorte configuration. Our temporal decay weighting modification is applied purely at inference time, replacing the single-text encoding step described in Section 3.1 with the per-experience encoding and weighted-averaging procedure described in Section 3.2. We sweep λ over {0.1, 0.2, ..., 1.0} in increments of 0.1, and additionally report the unmodified baseline for comparison.

### *4.3 Evaluation Metrics*

Following Decorte et al. (2023) and Senger et al. (2025), we report three metrics computed over the test set: Mean Reciprocal Rank (MRR), the average of the reciprocal rank of the ground-truth next occupation; and Recall@5

and Recall@10, the proportion of test instances for which the ground-truth occupation appears among the top-5 and top-10 predicted occupations, respectively.

## 5 Results

Table 1 reports MRR, Recall@5, and Recall@10 for the baseline and for selected values of λ; the omitted intermediate values follow the same trend described below.

| Metric | Base. | λ.1 | λ.3 | λ.5 | λ.7 | λ.9 | λ1.0 |
|---|---|---|---|---|---|---|---|
| **MRR** | .2543 | .2536 | .2520 | .2521 | .2570 | **.2581** | .2573 |
| **Recall@5** | .3552 | .3546 | .3568 | .3541 | .3602 | **.3607** | .3596 |
| **Recall@10** | .4256 | .4390 | .4401 | .4373 | **.4434** | .4401 | .4401 |

*Table 1: MRR, Recall@5, and Recall@10 for the baseline and for selected λ values. Bold/shaded cells mark the best value per metric. λ ∈{0.2, 0.4, 0.6, 0.8} (omitted for space) follow the same trend.*

Across the entire tested range $\lambda \in [0.1, 1.0]$, temporal decay weighting improves Recall@10 over the baseline (42.56%), with relative gains ranging from approximately +2.7% to +4.18%. The largest Recall@10 is obtained at $\lambda = 0.7$, reaching 44.34% , an absolute gain of 1.78 percentage points, equivalent to 32 additional test instances (out of 1,802) for which the correct occupation is recovered within the top 10.

For MRR and Recall@5, the trend is monotonically increasing up to $\lambda = 0.9$ before declining slightly at $\lambda = 1.0$. The best MRR is obtained at $\lambda = 0.9$ (25.81%, vs. 25.43% baseline, +1.49% relative), and the best Recall@5 is also obtained at $\lambda = 0.9$ (36.07%, vs. 35.52% baseline, +1.55% relative). At $\lambda = 0.9$, Recall@10 is 44.01%, still 1.45 percentage points above the baseline, though below the peak at $\lambda = 0.7$.

Recall@10 thus peaks at a moderate decay rate ($\lambda = 0.7$) and decreases slightly for larger $\lambda$, whereas MRR and Recall@5 continue to improve up to $\lambda = 0.9$. We discuss this divergence in Section 6.

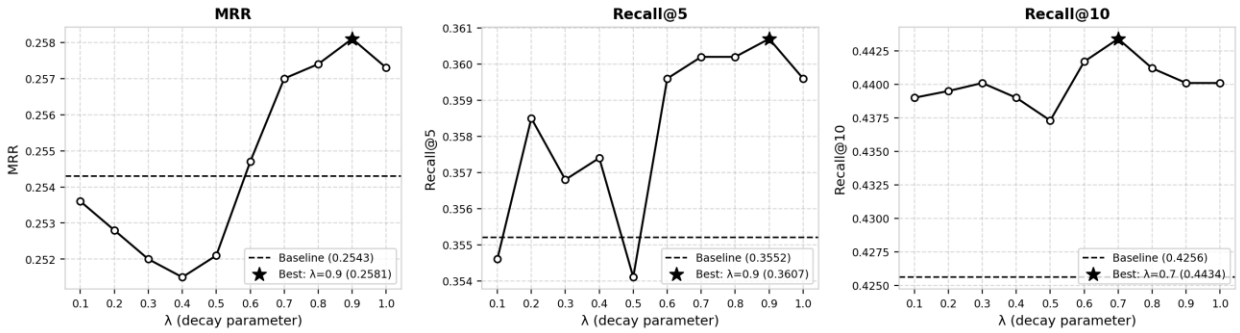


*Figure 3: Effect of lambda on MRR, Recall@5, and Recall@10. Dashed lines mark the baseline; stars mark the best lambda per metric.*

## 6 Discussion

### *6.1 Why does the optimal λ differ across metrics?*

At large $\lambda$ (e.g., 0.9), the weighted-average representation h is dominated almost entirely by the most recent job experience, since $\hat{w}(i)$ for $i = N-1$ approaches 1 as $\lambda$ grows. The resulting representation is therefore highly specific to the candidate's current role, producing sharper, more confident predictions , beneficial for MRR and Recall@5, which reward correctly ranking the target occupation near the very top of the list. At a moderate $\lambda$ (e.g., 0.7), the representation retains a meaningful contribution from earlier experiences as well, producing a representation that reflects a broader trajectory rather than only the current role. This broader representation is more likely to be similar to a wider set of plausible next occupations, which benefits Recall@10 , a coverage-oriented metric that only requires the correct answer to appear somewhere in a larger candidate set.

### *6.2 Implications for HR applications*

This precision–coverage trade-off maps naturally onto two distinct HR use cases. A configuration tuned for MRR/Recall@5 ($\lambda \approx 0.9$) is appropriate for tight shortlisting, e.g., matching a candidate to a small number of highly specific open positions where precision at the top of the list matters most. A configuration tuned for Recall@10 ($\lambda \approx 0.7$) is more appropriate for talent-pool construction in internal mobility planning, where the goal is to surface a broader set of plausible next roles for further human review, and a wider net is preferable to a narrow, highly precise one.

### *6.3 Practical advantages*

Because temporal decay weighting modifies only how the input representation to the (frozen) MLP is

constructed, it can be added to an existing embedding-based pipeline with a single additional step , per-experience encoding followed by a weighted average , without retraining any model component. This makes the modification inexpensive to deploy and easy to tune, since λ can be adjusted as a single scalar hyperparameter at inference time, even per use case (e.g., a smaller λ for talent-pool generation and a larger λ for shortlist generation within the same deployed system).

### *6.4 Limitations*

Our evaluation is restricted to the decorte configuration of KARRIEREWEGE; results on other configurations, covering broader occupational domains, remain to be verified. The decay parameter λ is a single global hyperparameter, shared across all candidates and career lengths; it does not adapt to individual career trajectories or to the number of experiences N. Finally, our method operates purely at the representation level and does not incorporate explicit skill-level information, which could provide complementary signal.

## 7 Conclusion and Future Work

We presented temporal decay weighting, a simple, inference-only modification to the MLP-based career path prediction pipeline of Senger et al. (2025). By encoding each job experience separately and combining the resulting embeddings with exponentially recency-weighted averaging, our method consistently improves over the baseline on the KARRIEREWEGE decorte configuration: Recall@10 improves from 42.56% to 44.34% (+4.18%) at $\lambda = 0.7$, and MRR and Recall@5 improve by +1.49% and +1.55% respectively at $\lambda = 0.9$ , all without retraining the underlying sentence-transformer or MLP components. We further showed that the choice of $\lambda$ corresponds to a precision–coverage trade-off with direct relevance to HR use cases such as shortlisting versus talent-pool construction.

Future work includes: (1) learning $\lambda$ directly from data, potentially as a per-candidate or per-position parameter; (2) evaluating temporal decay weighting on additional KARRIEREWEGE configurations and on other career-path and job-recommendation datasets; (3) combining temporal decay weighting with explicit skill-level representations derived from the ESCO taxonomy; and (4) studying adaptive decay functions that account for the total length of a candidate's career history.

## References


Matthew Bidwell. 2011. Paying more to get less: The effects of external hiring versus internal mobility. Administrative Science Quarterly, 56(3):369–407.

Marcus L. Call, Anthony J. Nyberg, Robert E. Ployhart, and Jeffery Weekley. 2025. On the move: The impact of internal mobility and internal comobility on unit- and organization-level outcomes. Academy of Management Journal.

Michael A. Campion, Lisa Cheraskin, and Michael J. Stevens. 1994. Career-related antecedents and outcomes of job rotation. Academy of Management Journal, 37(6):1518–1542.

Lu Chen, Nan Yang, and Philip S. Yu. 2022. Time lag aware sequential recommendation. In Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22).

Jens-Joris Decorte, Jeroen Van Hautte, Johannes Deleu, Chris Develder, and Thomas Demeester. 2023. Career path prediction using resume representation learning and skill-based matching. In RecSys in HR 2023: The 3rd Workshop on Recommender Systems for Human Resources.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of NAACL-HLT 2019, pages 4171–4186.

Ahmad Y. Hassan, Eman Fadel, and Nadine Akkari. 2022. Exponential decay function-based time-aware recommender system for e-commerce applications. International Journal of Advanced Computer Science and Applications, 13(10).

Jane Maley, et al. 2024. Talent management during digital transformation: Role of transformational leadership and resistance to change. Technology in Society.

Jiseong Oh and Sungzoon Cho. 2024. Measuring recency bias in sequential recommendation systems. arXiv preprint arXiv:2409.09722.

Christophe Ray. 2024. Internal mobility: A review and agenda for future research. Journal of Management, 50(1).

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using siamese BERT-networks. In Proceedings of EMNLP-IJCNLP 2019, pages 3982–3992.

David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. 1986. Learning representations by back-propagating errors. Nature, 323:533–536.

Elena Senger, Yuri Campbell, Rob van der Goot, and Barbara Plank. 2025. Toward more realistic career path prediction: Evaluation and methods. Frontiers in Big Data, 8.

Chen Zhu, et al. 2023. A comprehensive survey of artificial intelligence techniques for talent analytics. arXiv preprint arXiv:2307.03195.