Annabel Wei

AI, Machine Learning Engineer & Data Scientist
annabel.wei@yahoo.co.uk 078 288 33800 London, UK

Profile Summary

- Seasoned Machine Learning Specialist with an esteemed track record of executing complex projects, including an AI-powered podcast summarizing application and pioneering movie recommendation algorithms.
- Proficient in the complete lifecycle of AI & Machine Learning development, from data preprocessing to model evaluation.
- Demonstrated expertise in Natural Language Processing, Computer Vision, and Deep Learning, utilizing Python-based frameworks like TensorFlow, Keras, and Scikit-learn.
- Comprehensive experience in cloud platforms such as AWS and Alibaba Cloud, underscoring a solid foundation in AI & cloud infrastructures.
- Robust mathematical background, complemented by advanced knowledge in statistics and probability.
- Adept at harnessing the synergy of cross-functional teams to align with business goals, fortified by an M.Sc. in Artificial Intelligence. Passionate about converting intricate business challenges into actionable AI-driven solutions.
- Engaged collaborator with strong interpersonal skills, able to influence and align multiple teams with conflicting priorities to drive progress under aggressive time constraints.

Education

Birmingham City University M.Sc. in Artificial Intelligence Distinction **Birmingham, UK** 2022- 2023
Distinction **Coursework:** Computing for AI, Deep Learning, Applied AI, Impact of AI, Machine Learning, Data Visualization.
Dissertation Projects: Development of movie recommendation algorithms and machine learning models for revenue prediction.

London South Bank University B.A. (Hons) in Business Management 2.1 **London, UK** 2008- 2009

Technical Skills

Programming and Development: Python, JavaScript, HTML, CSS
Machine Learning and Deep Learning Frameworks: Scikit-learn, TensorFlow, Keras, Pyro
Data Manipulation and Visualization: Pandas, Matplotlib, Seaborn
Databases and Data Storage: MongoDB, MySQL
NLP & Computer Vision: Large Language Models, Speech to Text, PCA

AI & Machine Learning, Data Science Projects Experience

AI Powered App for podcast summarizing 2023

- Developed an AI-powered application within a team using Large Language Models and Streamlit to generate personalized weekly newsletters summarizing podcast episodes, assisting listeners in selecting relevant episodes based on guests, topics, and highlights.
- Designed a three-component architecture to extract and transcribe podcast content, convert the function to a backend, and develop a frontend for user experience using Streamlit.
- Incorporated Large Language Model (LLM) and Speech to Text model to optimize information extraction.
- Leveraged Modal Labs as the deployment service, ensuring efficient podcast discovery for listeners.

Movie Recommendation Systems 2023

- Pioneered AI-centric recommendation algorithms, deploying techniques like KNN-based Collaborative Filtering and advanced Bayesian Machine Learning, to craft highly personalized movie suggestions through data-driven insights.
- Handled extensive data preprocessing tasks, including missing value imputation, outlier detection, and feature engineering, on a dataset comprising various movie attributes and user ratings.
- Implemented Probabilistic Matrix Factorization using Pyro, employing Bayesian methods to provide a probabilistic framework, and utilized Stochastic Variational Inference for scalable and effective model training.
- Conducted comprehensive model evaluation, achieving an MSE of 0.85, MAE of 0.65, RMSE of 0.92, and R^2 of 0.78, to diligently select the most predictive model and ensure precise movie recommendations.
- Optimized models via hyperparameter tuning, particularly focusing on the ‘number of neighbors’ in KNN and Neighborhood-based models to enhance recommendation quality.
- Employed a robust data cleaning and exploration process, managing datasets with up to 45,466 entries across multiple files, ensuring coherent, unified, and clean data for model training and evaluation.

ML Regression Model Development for Movie Revenue Prediction 2023

- Developed and tuned multiple regression models, including Random Forest, Support Vector Regression (SVR), and Ridge Regression, for predicting movie revenues based on various features.

- Implemented XGBoost, a gradient boosting framework, to further enhance predictive performance and achieved notable improvement in R-squared and error metrics over traditional models.
- Achieved an R-squared value of 0.6873 using the Random Forest Regressor on the test dataset, indicating approximately 69% explanatory power on movie revenue variance.
- Enhanced the SVR model through hyperparameter tuning, elevating the R-squared value from 0.5526 to 0.5954 on the validation set and achieving an R-squared of 0.5834 on the test set.
- Employed and optimized a Ridge Regression model, obtaining an R-squared value of 0.5928 on the test dataset, elucidating 59.28% of the variance in movie revenues.
- Validated models' efficacy and generalizability using separate datasets, achieving key metrics like MAE of 0.3208, MSE of 0.3165, RMSE of 0.5625, and an R-squared of 0.6873 with the Random Forest Regressor on the test data.
- Explored opportunities for further enhancement by investigating alternative regression models and additional feature engineering to improve predictive accuracy.

NLP Sentiment Analysis Prediction 2023

- Developed and trained a sentiment analysis model for tweets, leveraging Support Vector Machines (SVM) and Decision Trees, through rigorous data preprocessing, feature extraction, and hyperparameter optimization, achieving an impressive accuracy of 87.8% with the SVM model.
- Employed advanced NLP techniques such as Tweet Tokenization, Spacy lemmatization, and removal of non-textual elements, alongside effective usage of bag-of-words, TF-IDF, and n-grams, leading to the creation of robust models.

Imagine Recognition with Computer Version for Cats & Dogs Prediction 2023

- Utilized Python and various feature extraction techniques to design image classification models, improving data quality and pre-processing speed, and achieving a classification accuracy of ~72% with the PCA model.
- Applied the PCA model to a new set of images and disseminated the findings and methodology through a comprehensive blog post and Python notebooks on Google Colaboratory.

NLP Deep Learning for Poem Creation 2023

- Devised a sequential neural network model for generating poems in the style of William Blake, leveraging syntactical and semantic analysis techniques and achieving an exceptional BLEU score of 99.8%. Model structure encompassed an embedding layer, bidirectional LSTM layers, dropout layers, and a dense layer with SoftMax activation.
- Identified opportunities for model improvement by refining sequence length, reconsidering the use of bidirectional LSTM layer, and implementing alternative sampling methods such as Top-K or Nucleus.

Professional Experience

Ginger Nut Training Apprenticeship Coach **London, UK** Oct. 2021 - Present

- Spearheaded course development including Level 3 Digital Marketer and Content Producer achieving 90% retention, 15% increased user interaction, and accolades.
- Streamlined resource redesign and leveraged email marketing, web content, and social media, reducing launch time by 90% and boosting user interaction by 15%.

Pearson Independent Assessor (Product Management, Online Courses) **London, UK** Sep. 2020 - Sep. 2021

- Orchestrated end-to-end assessment planning, mapping, and tracking for Pearson, ensuring alignment with quality management standards and metrics, thereby achieving a record of 100% timely student assessment completion.
- Pioneered the product vision for Pearson's suite of assessment tools, incorporating user research, market trends, and competitive analysis to ensure a customer-centric approach.

Arch Apprentices Learning & Development Specialist **London, UK** Feb. 2015 - Aug. 2020

- Mentored and trained 200+ apprentices in digital marketing, content production, and data analysis, collaborating with multiple firms such as Omnicom, StarCom, Publicis, Lloyds TSB, and Google, achieving a 100% student pass rate.
- Conducted over 400 customer interviews and data analysis, identifying learner pain points and developing relevant solutions, leading to exceptional and positive learning experiences.

IGNL Group Digital Marketing Manager **London, UK** Sep. 2009 - Jan. 2015

- Analyzed CRM strategies and Customer Journey Mapping touchpoints, managing the online store's performance and analytics reporting across platforms like social media, corporate website, and SQL database.

Certifications

Chartered Institute of Marketing CIM Level 4 Award in Digital Marketing - Google Level 4 Square Online - Google Individual Qualification and Analytics - IBM Data Science 101 & Python for Data Science - CoRise: Crash Course in SQL and Python - British Computer Society (BCS) Level 3 Principle of Coding - BCS Level 3 Marketing Principles and Digital Marketing - AWS Machine Learning Foundations 2022 - Alibaba Cloud Practitioner, Pendo Product Management - Pendo Product Analytics