Fan Yeliang

leo-van@hotmail.com (+86) 186-3159-3756 https://leovan.me

Education

Hebei University of Technology (HEBUT), Tianjin, China *2012 – 2015*

M.S. in Information Management (GPA: 85/100 Top 2%)

- 2nd Prize Scholarship (Top 5%), Outstanding Student (Top 5%), Outstanding Student Leader (Top 5%)
- Outstanding Graduate Student of Hebei Province (TOP 2%)
- 1st Prize of ChuangQingChun Business Plan Competition of Hebei Province (Team Leader)

Hebei University of Technology (HEBUT), Tianjin, China *2008 – 2012*

B.S. in Business Administration (GPA: 87/100 Top 2%)

- 1st Prize Scholarship (Top 2%), Outstanding Student (Top 5%), Outstanding Student Leader (Top 5%)
- President of Student Union of School of Management

Work Experience

Meituan *Risk Data Mining Expert* *Aug. 2020 – Present*

Security & Risk *Risk Data Mining Expert* *Aug. 2020 – Present*

JD *Senior Algorithm Engineer* *Apr. 2015 – Aug. 2020*

- High-potential Employees of JD Group (2019)
- Honorary lecturer of JD Digits (JD Finance) (2017, 2018)
- Bronze medal certified lecturer of JD Group (2019)
- Innovation Seed award of JingYa Cup Innovation Competition in JD Group (2017, All Seeing Eyes project, ranking **20/378**)

Intelligent Agriculture *Algorithm & Product Leader* *Jun. 2019 – Aug. 2020*

- **Intelligent Farming and Intelligent Poultry Solutions**: Lead the conception and design of intelligent farming and intelligent poultry business models and technology solutions. Led algorithm and product team members to build the data and algorithm models, and design the prototype of SaaS and APP **from 0 to 1**, which achieves the full solution from MVP to the real production environment.
- **Intelligent Environmental Control**: Design and development of intelligent environmental control algorithms and solutions based on time series analysis, deep learning and reinforcement learning. The intelligent environmental control algorithm consists two parts: environment models and control models. It realizes the reuse of models for same crop and livestock in different environments. With the expert knowledge engine and machine learning algorithms, the control error of various environmental indicators is reduced by **50%+** compared with the farmers while keep the regular yield of vegatable, and total average cost (including water, electricity, fertilizer and etc.) is reduced by **20%+**. In the 24-hour Hackathon simulation challenge of the 2019 International Autonomous Greenhouse Challenge, we achieved **4/21** of artificial intelligence strategy methods and **9/21** of net profit in virtual tomato planting.
- **Intelligent Eggs Collection**: Design and development of an intelligent eggs collection device and algorithm based on computer vision and sensors. During the eggs collection process in cage raising mode, with the data collected by cameras and sensors, it achieves the eggs counting and the belonged cages identification with accuracy of **99%+**. Through the belonged cages identification, it is possible to analyze

egg-feed ratio accurately, providing strong data support for hens elimination and a more detailed data source for eggs traceability.

Daat (Complex Network & Knowledge Graph) *Project Leader* *Apr. 2018 – Jun. 2019*

- **Data Knowledge Engineering & Data QA System**: Design and development of ontology of data warehouse, data market and data tools. Based on the ontology and the extracted knowledge, we build the knowledge base of data. We also develop the data QA system with techniques such as: intent classification, slot filling, query rewrite, ranking and question matching based on DSSM. Data QA system is aimed at improving the usability and convenience for users to make use of data warehouse and data market. It can also answer the questions related to data concept, data processing flow and data tools. The system serves **3000+** internal users, and manual service with data related problems is reduced by **50%+** with its help.
- **Automatic Sensitive Information Identification**: Development of automatic sensitive information identification for data warehouse, which helps to make data encryption policy. The model is based on the Wide & Deep network with meta-information of the data (e.g., table name, table comment, column name, column comment, etc.) and value-information of the data (e.g., the data values of every column). Building the Wide network with extracted traditional features and the Deep network with text features using Char Embedding + CNN, it achieves **95%+ of the F1-Score** on test data.
- **Large Scale Heterogeneous Network Embedding**: Development of large scale (**ten millions of vertexes and hundred millions of edges**) heterogeneous network embedding algorithm. We implement the algorithm based on meta-path with rich business meanings, and provide the embedding results as features for other business models, including risk management, marketing and recommendation.
- **Recommendation and Marketing based on User Network and User Behavior**: Leveraging historical orders, we build a large heterogeneous network of users which contains users, address, goods, and etc. With the embedding results of this network, we develop an algorithm for candidates generation of recommendation, which achieves **20%+ improvement** compared with traditional methods.

All Seeing Eyes (Chinese Address Analytics) *Project Leader* *Apr. 2015 – Apr. 2018*

- Development of **Chinese address analytics algorithms**, including: segmentation, classification, integrity, POI identification and similarity **accuracy 90%+)**.
- Development of **Address Profile System** based on the basic algorithm engine. It increased the conversion rate of users by 30%+ in the offline payment service.
- Development of the anti-fraud and credit model based on the Chinese address analysis system. The anti-fraud model identified illegal encashment orders with **200,000 CNY/day**, and more than **10 million users** were granted credit with the credit model.
- Development of **Enterprise Address Profile System** based on the basic algorithm engine which serves for internal and external users with offline data and realtime query service on **JD Enterprise Credit** (`https://icredit.jd.com`).
- Development of **Rural Finance Service Station Location Selection** based on the Address Profile System. It provides decision support for offline rural finance service station selection.

User Behavior Analytics *Algorithms Engineer* *Oct. 2017 – Dec. 2017*

- Development of a user behavior representation method named on Behavior2Vec. Based on hierarchical clustering and depth search, a hybrid model for identifying user abnormal behavior is proposed. Compared with Bag of Words and N-GRAM methods, the number of abnormal users identified is **3+ times** of traditional methods.

Mortgage Loan *Algorithms Engineer* *Apr. 2015 – Oct. 2015*

- Development of a hybrid product life cycle identification model based on Bass Diffusion model, optimized time series similarity method and clustering method. It got an **accuracy of 95%+** when identifying the excess inventory products, which helped to make loans goods pledge decisions and calculate the loan-to-value ratio.
- Development of product information fusion model and system with ElasticSearch which got **90%+ recognition accuracy** and provided accurate and relevant information, such as price, etc.

Skills

- Programing: Python: ★★★★☆ R: ★★★★☆ SQL: ★★★★☆ JavaScript/TypeScript: ★★★☆☆
- Frameworks: Qt: ★★★☆☆ PyTorch: ★★★☆☆ Tensorflow: ★★★☆☆ React: ★★☆☆☆
- Tools: Axure: ★★★★☆ Sketch: ★★★★☆ Omnigraffle: ★★★★☆
- Language Skills: CET-6: 518, fluent in communication, reading and writing skillful in English.

Papers

- Zhou, F., Yin, H., Zhan, L., Li, H., **Fan, Y.**, & Jiang, L. (2018). A Novel Ensemble Strategy Combining Gradient Boosted Decision Trees and Factorization Machine Based Neural Network for Clicks Prediction. In *2018 International Conference on Big Data and Artificial Intelligence (BDAI)* (pp. 29-33). IEEE.
- Li, J., **Fan, Y.***, Xu, Y., & Feng, H. (2013). An Improved Forecasting Algorithm for Spare Parts of Short Life Cycle Products Based on EMD-SVM. In *Information Science and Cloud Computing Companion (ISCC-C), 2013 International Conference on* (pp. 722-727). IEEE.
- **Fan, Y.**, Li, J., Chu, C. (2014). IEAF: A Hybrid Method for Forecasting Short Life Cycle Spare Parts. *Unpublished.*

Patents

- A kind of Chinese address segmenting method and system (CN 105159949) 2015
- Product inventory predicting method and product inventory predicting device (CN 106056239) 2016
- Data warehouse information processing method, device, system, medium (CN 109388637) 2018
- A kind of data processing method, device, equipment and medium (CN 110309235) 2019
- Method and apparatus for generating information (CN 112395490) 2019

Open Source Projects

- Github: `https://github.com/leovan` Followers: 208
- Data Science Introduction with Python: a tutorial of data science based on Python. Stars: 57
- Data Science Introduction with R: a tutorial of data science based on R. Stars: 98
- Sci-Hub EVA: a cross-platform Sci-Hub GUI application. Stars: 878
- Hive Functions: useful custom Hive functions. Stars: 3
- Cytoscape Manual: Cytoscape manual (Chinese Version). Stars: 11