Planning and development information system based on AI recommendation algorithm

I. Project Basic Information

1. Application Industries & Scenarios

The IPM system carries out career or academic further study planning for users at different stages and matches individual development plans.

The first phase of the IPM system is to match different grades of college students with target positions or majors that match their personal backgrounds, and to provide training programs for career development or further study development planning. With the accumulation of data and the development of business, the target group in the future will gradually increase the roles of high school students, in-service personnel and so on.

2. Core AI Technology

□Computer Vision	□Voice Recognition	■NLP	■Deep Learning	■AI Decision
□Federal Learning	□Knowledge Graph	□AI Robotics	□Smart Chip	□HPC
□Multi-modal Machine Learning		■Others (Please provide details): Machine Learning		

3. Application Development Stage

■Hypothesis Validation	□Alpha	□Commercial Implementation	□Growth

II. Project Content

1. Background and Significance

At present, the number of fresh undergraduate graduates in China has reached more than 4 million per year. According to the survey data, the undergraduates who are studying are more confused about continuing their studies or seeking job opportunities after graduation. The issues we need to focus on are how to make college students maintain a clear life plan for the future, and how to make colleges pay more attention to the core skills required by the industry. In addition, most internet companies spend more and more time cultivating management trainees. If they use this platform in advance, it will not only make students more competitive in job hunting, but also reduce the company's training costs for fresh graduates.

Based on the above pain points, we conducted an online questionnaire survey on some undergraduates in Beijing in the past month and obtained a total of 1,169 questionnaires. Among them, the ratio of undergraduates who hope to pursue further study and find a job after graduation is 6:4. According to the survey, 18% of the undergraduates are clear about their future development intentions, but they are unable to determine what means to use to achieve the set goals. There are 31% of students who are unable to match the position accurately, because they do not understand the necessary skills required for the job, so that they cannot exert their highest level in the job interview, which indirectly wastes the company's HR resources. Additionally, about 28% of students choose majors that are not of their own interests, resulting in a low degree of matching between future job searches and majors, which indirectly wastes the educational resources of colleges and universities. Finally, 23% of the students look ahead and cannot judge whether they should choose a job opportunity or pursue a higher degree, which may make undergraduates waste the most precious time in their lives.

In conclusion, all the conditions leave the global unemployment rate projected to remain above the 2019 level of 5.4% until at least 2023 (data from the 2022 UN Sustainable Development Goals Report). In addition, some fresh graduates may change jobs frequently later because they work in occupations and positions that do not match or are not interested in. This will make these students entering the workplace full of anxiety about the future and will not be able to realize their social value in the short term. Based on the above considerations, we believe that it is necessary to develop a digital and intelligent information system that can save college students' information retrieval time and provide students with clear development plans and training courses. We hope that this system can reduce the helplessness and anxiety of the students entering the society at the beginning. We hope that they can face future professional challenges with a confident and sunny attitude.

2. Application

In order to make college students have a sense of direction for future planning and shorten the information retrieval time for career/academic planning during their studies, out team members make great efforts to develop the *Planning and Development Recommendation Based on AI Recommendation Algorithm System (Intelligent Planning Master, IPM System)*. There will be a stage of confusion about the personal future plans whether who is a group of students (high school students, undergraduates, graduate students, and doctoral students) or professionals with many years of working experience. Based on these scenarios, the IPM system will provide them with tailored-made career or further education development paths.

The core function of this system is the data-intelligence strategy recommendation algorithm platform of the underlying architecture. Specifically, the IPM system is based on the user's function selection. The background of the system can automatically judge and select the required algorithm according to the users' needs and provide solutions such as classification, prediction, and so on. For example, according to the user's own background input, the IPM system can evaluate the background score, calculate the matching degree of the development path, and output the training plan, which will improve the quality of education, and complete the path achievement analysis. The specific algorithm and function application are as follows.

- Use crawlers to extract keywords from job search platforms and college ranking platforms, and then push the industry's top majors and top job requirements in real time through phone, email, and web.
- Use decision tree algorithm to match multiple optimal careers/professionals for users. This algorithm requires less data for training, and the model can be manually pruned to achieve higher accuracy.
- Use the collaborative filtering algorithm of items to recommend the most interesting jobs or majors for users. Among them, users with similar interests, training courses, and development directions are also relatively similar. In addition, this algorithm can also increase the recommendation of students' course learning scope, which make the recommendation scheme more diverse.
- Use classification algorithms to provide users with clear development plans and training courses. After the user's background input and probability statistics calculation, the IPM system will customize different algorithms and output classification results, such as Naive Bayes, Decision Tree, Neural Network, etc. These algorithms can be applied to large databases with simple methods, high classification accuracy and high speed.
- Use the user-based collaborative filtering algorithm to match students with similar development plans and communicate with each other. This algorithm recommends based on

user similarity, so it has stronger social characteristics.

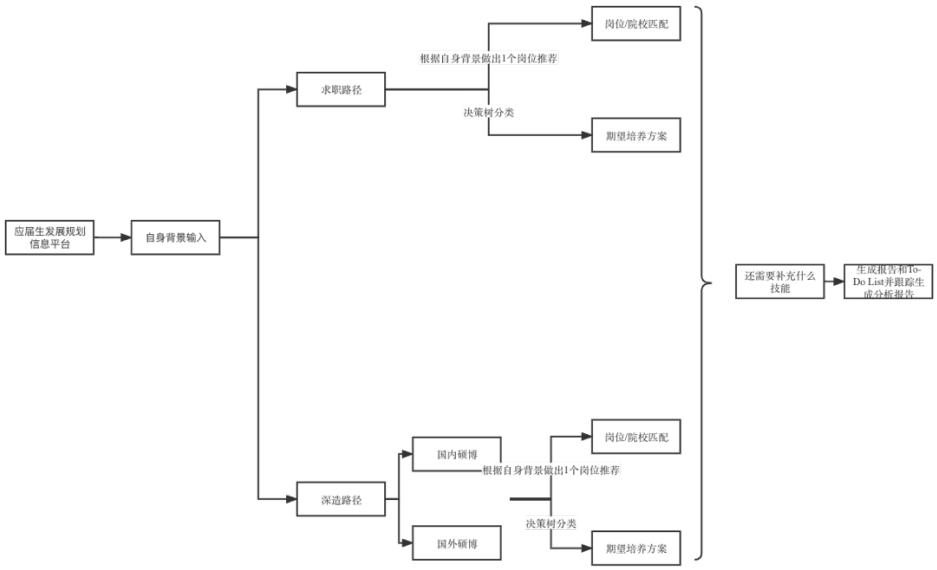
- Use the k-mean clustering algorithm to match the alumni of the corresponding HR and the destination college, which can communicate with each other online or by email.
- Use the collaborative combination of multiple AI algorithms to provide periodic feedback and analysis reports on the achievement of users' training plans, and output more intuitive visual reports.
- Provide forum discussion, recommendation of alumni resources, recommendation of HR resources, and services such as AI assistant Q&A and industry resource Q&A.

3. Implementation Method

3.1 Design Concept

The IPM system aims to use the AI algorithm that can automatically judge in the background to improve the efficiency of information screening for users in different stages of study or job hunting, and to provide accurate and effective development plans.

Nowadays is the information age. While it is convenient to obtain information on the Internet, there is still a lot of redundant information, which makes it impossible for many people to screen effective information, resulting in missing many jobs search/further education opportunities. For example, for undergraduates, four years of college are particularly valuable. We hope to display the top majors and top job positions for users in real time, so that they can always have informational connections with the industry and universities. In addition, we hope to provide corresponding training programs, so that users can make plans for their future as soon as possible and be more competitive in the process of future employment or further study.



3.2 Core Technology Route

3.2.1 AI Decision-Decision Tree Algorithm: User positioning

The IPM system uses a decision tree algorithm to classify the user's background, which can clear its own position.

Based on several recruitment websites, QS Ranking, and US News Ranking websites, we accumulate professional and job-related labels and data, and classify different user portraits according to the corresponding labels, which are used as a powerful training data set in the IPM system. The user enters their own background information, and the IPM system calculates the matching degree of the corresponding position or higher education institution according to the user background and the information entropy of the system data, which can select the corresponding planning and development plan.

Based on various positions and college majors in the market, the IPM system stores a large number of capability labels for matching. These data can calculate the matching degree for the user, provide the user with the skills improvement required to achieve the target position and target college, and match the corresponding plan and upgrade courses.

3.2.2 NLP-Text Analysis: Recommendations for User Skill Improvement

The IPM System uses NLP-Text Analysis for users' screening to conduct real-time calculation and recommendation of employment guidance and college recommendation functions.

The IPM system uses crawlers to extract keywords from job descriptions on recruitment websites, directly locate and strengthen skills, and directly screen target domestic and foreign institutions according to their needs. The IPM system creates a skill tag word cloud based on the frequency of keywords, so that users can see the required skills more intuitively. Positions and professional representatives with high information entropy have a large market demand, which can show users top occupations and professions, so that they can have a more sense of direction for future development.

3.2.3 Machine Learning-Clustering Algorithm: User Planning Development Path Recommendation

The IPM System uses the Machine Learning-Clustering Algorithm to recommend the improvement plan for the planned development path expected by the user.

The IPM system stores a large number of training data sets, such as the label of career planning path corresponding to "job positions - skill labels - enhancement courses - online business projects", and the label of further study development path corresponding to "college majors - skill labels - enhancement courses - scientific research projects". If the user cannot provide their own ability label, the IPM system will ignore their own background temporarily, then input the user's desired development path as data input. Finally, the IPM system will recommend a development plan according to the user's choice for the user's final confirmation.

Users can mark "Recommended" or "Not Recommended" for the recommended development plan. If the recommended path is different from their own choice, the students' mark can assist in maintaining whether it matches, which is beneficial for the system to do deep learning.

3.3 System Architecture

3.3.1 Construction of The Underlying Algorithm Platform

The IPM system implements various algorithm requirements through a set of algorithm platforms built at the bottom. At present, the platform capability is to manually select the optimal algorithm model and complete the automatic model training.

(1) Data upload web display

Users input their own background information, store it on the web platform and display it. The system saves the data in the background and conducts pre-analysis of each planning.

(2) Manual selection algorithm

In the initial stage, users are required to manually select algorithms according to their own planning and development needs. This is not only the collection of training data for the IPM system, but also the most direct planning algorithm results for users.

(3) Model evaluation prediction

Although the user is required to define the type of AI recommendation algorithm required by means of manual selection, the IPM system can also provide evaluation guidance and prediction according to the user's choice. This procedure will guide the user to choose the AI recommendation algorithm model with a higher degree of matching, which can improve the user's satisfaction with the output of the planning and development plans.

3.3.2 System Functions

(1) Automatic data marking

The IPM system can automate data acquisition and convert and mark different data types according to certain rule. For example, type variables such as gender, ordinal variables such as multiple educational backgrounds, and numerical variables such as age (with implicit meaning, such as length of service).

(2) Automatic data correction

The IPM system can automatically identify and filter problematic data to ensure the integrity of data verification. In addition, the IPM system can automatically roll back and correct historical data over time to ensure the validity of data verification. For example, the previous age samples need to be updated periodically.

(3) Graphical display of data label accuracy

The IPM system conducts data sampling by selecting sample test data, and performs hierarchical

display of charts through data marking, including categorical data, numerical data, time series data, etc. In addition, the IPM system can also perform correlation analysis on the sampling data.

(4) Automated data dimensionality reduction

The IPM system can perform correlation analysis through automatic transformation and dimensionality reduction of different dimensions of data. For example, in the TIME dimension, the prediction of time series data requires dimension reduction of the data. We can convert TIME to different dimensions through indicators, and then use multiple column dimensions to correspond.

3.3.3 Data Algorithm Selection

The data corrected by marking can automatically select the optimal algorithm through the platform.

(1) Conventional marking method adjusted by algorithm engineers

- Low accuracy: Adjust the optimization function automatically and the activation function.
- Centered accuracy: Data is under-fitted, and variables can be added based on transformations.
- Large correlation: Data overfitting, the model can increase the amount of data and regularization automatically.

However, the IPM system can adjust the parameters automatically. When the IPM system accumulates enough data in the background, the system can train the same batch in real time through different parameters of each set of input data, select the optimal solution from multiple algorithm results, and then complete the automatic adjustment of parameters.

(2) The IPM system can optimize the algorithm platform from the perspective of the algorithm, which will be more intelligent and advanced than the traditional algorithm platform. For a defined model, the IPM system can perform real-time synchronous training of multiple models to select the optimal result and invert the data model. For example, for different types of algorithm models, the IPM system will output various good and bad result schemes when training multiple parameters. However, with the increase of training time, the IPM system can obtain the optimal algorithm scheme from the results of each batch, and automatically select the optimal algorithm model according to the algorithm results.

3.4 The Deployment of Software and Hardware

3.4.1 The Deployment of Software

Install the Python environment and MySQL database through the official website. Download the corresponding version according to your computer needs. At the same time, use the terminal to install the Django framework and MySQL package that Python needs to call.

3.4.2 Hardware Requirements

- Database server: 200,000 tpmc minicomputer
- Algorithm server: 8C, 32G PC Server, GPU 3070/3080 * X card
- Web server: 8C, 32G PC Server

- Disk Array: grows according to actual access per year
- NAS: grows according to actual visits per year

3.4.3 Part Code Display of Core Function

（1）Home Complete Project Space Selection

```python
@app.route('/',methods=['POST','GET'])
def index():
    '''
    首页加载文件
    '''
    if request.method == 'POST':
        session['work_name'] = str(request.form.get('work_name') or 'default')
        work_name = session.get('work_name')
        text = '构建私有空间成功，请牢记名称：' + work_name
        return render_template('index.html', text=text, work_name=work_name)

    work_name = session.get('work_name')
    text = '当前工作空间(切换请清空缓存):' + str(work_name or '')
    return render_template('index.html', work_name=work_name, text=text)
```

（2）Load Model

```python
@app.route('/loadmode', methods=['POST', 'GET'])
def loadmode():
    '''
    加载模型 ： 加载model目录下数据模型
    '''
    if request.method == 'POST':
        model_name = request.values.get('model')
        session['model_name'] = model_name
        clf, alg = operation_model(operation = 'm_load')
        text = f'模型{model_name}, 算法{alg}加载成功..'
        return render_template('loadmode.html', text = text)
    else:
        ip_path = get_model_path()
        file_dict = get_file(ip_path)
        text = '请选择一个模型文件加载...'
    return render_template('loadmode.html', show_file_list = file_dict.keys(), text = text)
```

（3）Predict Model

```python
@app.route('/predict', methods=['POST', 'GET'])
def predict():
    '''
    预测模型
    '''
    text, desc = '',''
    model_name = session.get('model_name')
    if model_name is None:
        return redirect('/loadmode')
    clf, alg = operation_model(operation = 'm_load')
    desc = f'当前模型:{model_name}, 当前算法:{alg}'
    # 模型使用算法
    if request.method == 'POST':
            # 预测数据加载
            f = request.files['filename']
            df = pd.read_excel(f)
            x = df.values
            # 开始预测
            pre_res = clf.predict(x)
            pre_df_res = pd.DataFrame(pre_res, columns=['alg'])
            merge_df = pd.merge(df, pre_df_res, left_index=True, right_index=True)
            operation_model(df=merge_df, operation='df_pn_save')
            # 返回结果
            model_name = get_security('model_name', add_name='.model')
            desc = f'当前模型:{model_name}, 当前算法:{alg}'
            text = '''分类结果:{pre_res}'''.format(pre_res=str(pre_res), alg=alg)
    return render_template('predict.html', text=text, desc=desc)
```

3.5 Future Vision

At present, the main users of the IPM system are students at any stage from freshman to senior year. In the future, we hope to expand the number of users, and gradually extend it to not only provide high school students with undergraduate major selection and future development plan guidance, but also provide promotion or career change plan guidance for in-service employees at the working stage. The future user portraits are mainly as follows.

(1) High school students

Through the historical background (CTFs, programming competitions, or hackathons. Peer-reviewed research, significant awards, significant community service, participated in sports or joined a performance group) of high school students, we provide high school students with options for future development of undergraduate majors or overseas school projects.

(2) Employees

If employees want to improve your chances of promotion in this industry, the IPM system can provide employees with cooperative online group business projects, technical programming competitions, etc., which can not only make more connections but also improve their comprehensive ability. If employees want to change careers, the IPM system can also provide cooperative online courses, online micro-degree and degree courses for training according to employees' intention. At the same time, the IPM system provide employees with internal promotion opportunities from different Internet companies to achieve the purpose of career change.

(3) Business going overseas

When the IPM system has accumulated enough data and analysis reports in China, we hope to promote the IPM system to overseas users and cooperate with more overseas universities, overseas Internet companies, and start-up companies in various fields. We hope to expand our career development platform business. We hope to expand our career development platform business to meet more people resources. This vision can not only cultivate more excellent all-round "H-type" talents for the society, but also match talents for enterprises to improve the overall economic level of the world, thereby promoting the sustainable development of the world.

4. Key Pillar

4.1 Scenario Innovation

4.1.1 Optional Job Search or Further Study Path

Nowadays, there are many recruitment websites with complicated information. There are also many online courses in cooperation with major universities, and training courses about SDE and DS for job hunting. However, the IPM system integrates all the above resources, which can give users a training

plan after processing and analysis by the multi-dimensional AI recommendation algorithms.

4.1.2 Job Search Path

Most recruitment websites only have job descriptions and background matching scores. However, the IPM system will integrate keywords and provide clear job search development paths, skill training courses, etc. The IPM system extracts keywords for the relevant job descriptions through web crawlers, which not only enables students to have a subjective perception of core skills, but also allows students to have a more sense of direction in undergraduate courses and extracurricular studies, making them more likely to be successful in job hunting.

4.1.3 Further Study Path

Most online education websites only provide courses according to subject categories and students need to select and study by themselves. However, the IPM system can recommend scientific research projects or professional courses according to the needs of students.

4.2 Function Innovation

At present, the network resources are relatively scattered, and students need to manually go to various websites to search for information, screen and determine how to improve their own abilities. However, the IPM system will automatically determine the matching position/major according to the background and needs of the students, then output the training plan and To-do List. Finally, students can mark completed items and generate weekly achievement report analysis.

(1) There are career paths and further study paths to choose from.

(2) Give the best position/best professional match according to the own background.

(3) Real-time updates of popular majors/popular positions, allowing users to keep abreast of the current social situation.

(4) The IPM system can recommend development plans according to the user's background and the input direction of interest. The solutions come in many forms and can be customized. The IPM system can calculate the matching degree and output the recommended training plan according to the combination of the determined occupational position/college major and its own background. Moreover, the IPM system can match the best jobs/professionals based on skill tags to deliver resumes in batches.

(5) The users set the time required to achieve the goal through the IPM system. The IPM system will recommend courses corresponding to the class hours according to the user's choice.

(6) Users can set their own graduation date or entry date, etc. The IPM system will generate weekly/monthly achievement reports to track the achievement progress of user plan development.

4.3 Applied Technology Innovation

The IPM system can train accurate models in the background. The application technology advantage

of the IPM system is that the algorithm is manually selected according to the user's needs in the early stage. The users will receive the results after the model is calculated. In the later stage, the IPM system can automatically optimize the subsequent algorithm and scheme output as time accumulates.

4.3.1 The IPM system can optimize the algorithm platform from the algorithmic point of view.

4.3.2 Automated Data Marking and Correction

(1) Different data types are converted, marked, and filtered according to certain rules to ensure the integrity of data verification.

(2) With the passage of time, the historical data is automatically rolled back and corrected to ensure the validity of data verification.

(3) Carry out hierarchical display and correlation analysis of charts through data sampling and data marking.

(4) Perform correlation analysis through automatic transformation and dimensionality reduction of different dimensions of the data.

4.3.3 Automated Selection of AI Algorithms

The IPM system performs multi-model training based on the user's background input. Finally, it selects the optimal result and gives users the optimal development plan by selecting the data model automatically.

4.4 Solution Implement Ability

4.4.1 Server Cost

The correspondence between skill labels and employment positions and further study majors comes from real-time crawler data from major recruitment websites and college admissions websites, which can ensure the accuracy of student information recommendation. The main functions of the server are simple queries, complex queries, algorithm calls and data interaction, etc.

(1) The bops required by the business =
peak business [minute] / 60 * response time * complexity factor / server load rate

The meaning of each parameter:

- 60: 60 seconds
- Business complexity factor: describe the number of standard bops transactions corresponding to a business transaction
- Service response time: describe the response time of a service under peak conditions, calculated as 120s
- Server Load Rate: describes the load on the server running at full, calculated as 70%

(2) Business peak per minute =
peak business volume [pages/day] * 80% / (working time [minutes] * 20%)

The meaning of each parameter:

- 80%: the amount of business during peak hours
- 20%: concentrated working hours for business volume
- These two parameters mean that 80% of the business volume during the peak period is mainly concentrated in 20% of the working hours

4.4.2 Big data/Algorithm Platform Cost

(1) Network requirements

For daily user visits, monitor according to the distribution of users at all levels (country/province/prefecture/county), which can interact with the cooperative unit system and interconnect with the external network institution system.

(2) Data flow analysis

The main network traffic in the system includes business volume statistics per unit time. The network bandwidth required in the execution of each business needs to take into account the actual network bandwidth required for concurrency and network utilization efficiency (the efficiency is about 75%) and the average usage of CPU and memory required during daily and peak runtimes.

4.4.3 Cooperation Cost of Enterprises and Institutions

(1) In the future, we will cooperate with the school admissions office, HR in the companies, and online courses to refine our label classification and system resources.

(2) In the future, we will cooperate with major job search websites to obtain job description information to improve the accuracy of system recommendation and save the labor cost of interviewers.

(3) If we cooperate with universities and provide training programs. Users can take courses from their own schools. At the same time, the IPM system also guides the future training programs of colleges and universities. They can update the syllabus according to the hot skills of each post, which can be closer to the needs of social talents

(4) The cost of other front-end development engineers, promotion, sales and so on.

III. Application Analysis

1. Market Analysis

Nowadays, the number of college students is on the rise every year. Under the premise that the admission rate of postgraduates remains unchanged, the number of freshmen applying for postgraduate examinations is also increasing. At the same time, the three-party contract rate of fresh graduates who choose the job search path has declined. From the above data, the IPM system can help students understand market demands and make career or further study plans in advance.

With the popularization and enhancement of informatization, undergraduates have stronger and stronger ability to obtain information. The average academic level of fresh graduates is also increasing every year, and more and more students hope to achieve greater success in the future by improving their abilities as soon as possible. Therefore, the IPM system accurately locates the pain points of undergraduates, which also greatly improves the acceptance of the market.

In the future, we hope to cooperate with major universities and Internet companies to form a national university student development planning network, so that the solutions will become more sustainable and practical.

2. Business Model Analysis

2.1 Market positioning

At present, we provide employment and domestic and foreign development path recommendations for undergraduates at any stage from freshman to senior year.

2.2 Potential Competitors

From a solution point of view, there is currently a gap in the market. Our main users are undergraduate students, and the main use stage is the development planning stage after undergraduate studies and before the recruitment website

For overseas study paths, consulting companies studying abroad are potential competitors. They will provide programs for students studying abroad according to different grades and backgrounds, so as to improve the success rate of study abroad applications. But in comparison, our system will be smarter, save labor costs, and our business scenarios will be wider.

For overseas study paths, consulting companies in the field of overseas study are potential competitors. They will provide programs for students who are willing to study abroad according to different grades and backgrounds, so as to improve the success rate of study abroad applications. By contrast, our system will be smarter and save more labor costs. Moreover, our business scenarios are wider.

2.3 Cooperation Mode

(1) Cooperate with HR systems of major companies and recruitment platforms in the industry. They can get profit sharing if users join the company.

(2) Cooperate with online courses of domestic and foreign colleges and universities. They can obtain profit sharing according to different businesses after the recommendation generates course traffic or the students pay for the class.

(3) Cooperate with industry career development consulting companies and use the consulting companies' information sharing on industry career development and academic planning to improve the recommendation algorithm of this system.

(4) Finally, the IPM system gradually charges membership fees to promote the membership system, so that users can enjoy more diversified personalized services.

2.4 Market Development Strategy

Starting from undergraduates, the user group with the largest volume and relatively more confused users, the corresponding big data is accumulated. With the accumulation of big data and the expansion of business, we will provide more development plans, such as scientific research, competition training and other learning programs. In the future, we will expand to more user groups, so that high school students and incumbents around the world can also enjoy the recommended solutions brought by the IPM system.

IV. Social Benefits

By using the IPM system, target users can clear up their confusion and worries about future development faster. By implementing the recommended development plan, the target users can actually learn the corresponding vocational skills. For career-oriented users, it can improve the employment rate of undergraduates after graduation and the signing rate of tripartite contracts. For further education-oriented users, it can improve their enrollment rate of master and doctoral degree. From a longer-term perspective, a suitable and interesting development plan can not only generate greater driving force for users to improve themselves and serve the society, but also provide the society with more talents that are highly matched to the jobs. These benefits can improve the overall economic benefits of the society.

Moreover, the effective application of the IPM system can not only improve the overall skills and quality of new workplace groups, but also improve the employment rate and job matching rate of fresh graduates from the perspective of social services. With the increase of business scenarios and user groups, the IPM system can integrate global social culture, make users' development plans international and diversified, and allow more people to have a broader vision.

Appendix

1. Projects' (Products, solutions, etc.) products or actual application pictures (If any).